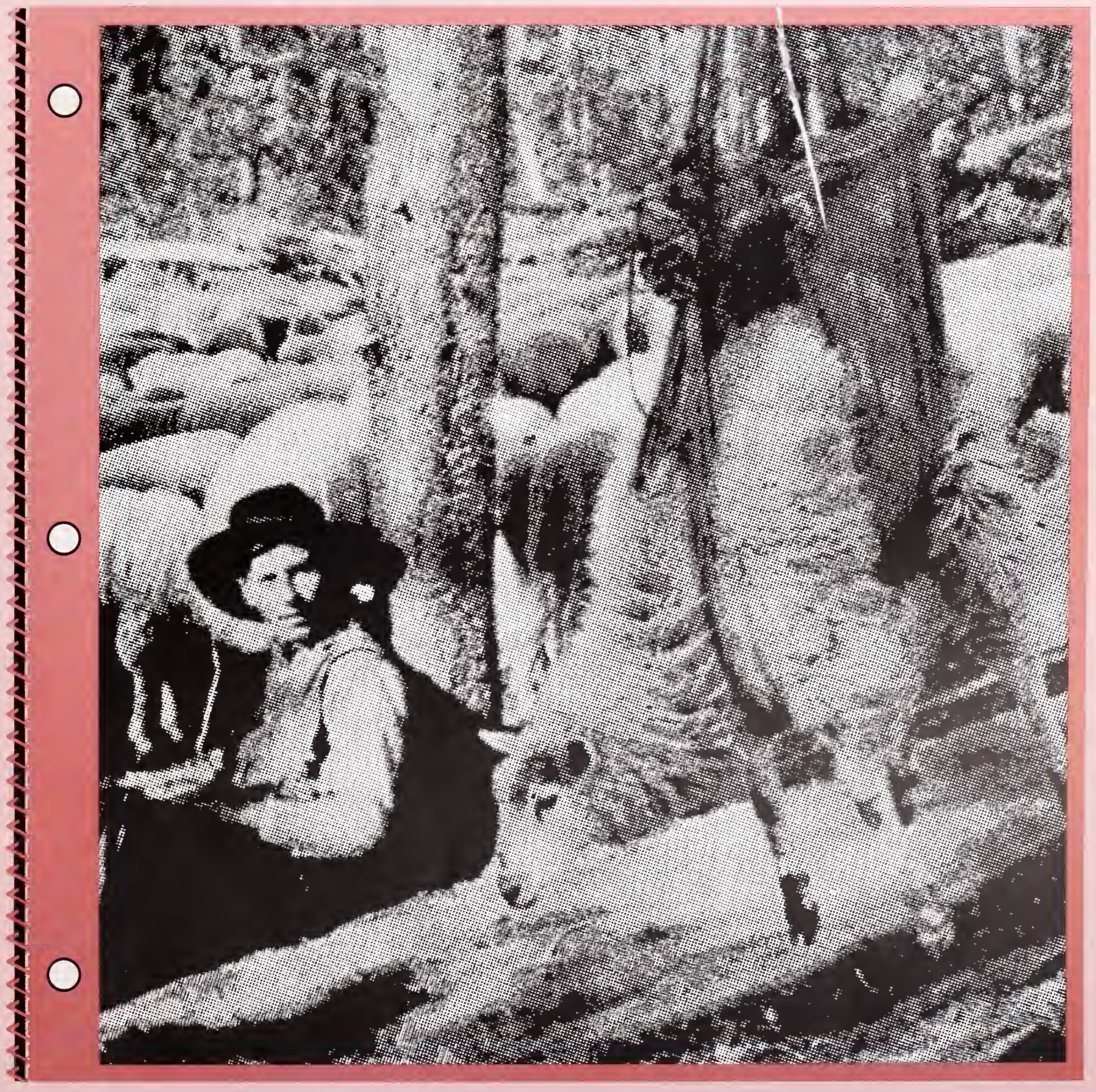
United States Department of Agriculture

A Primer on Evaluation and Use of Natural Resource Information for Corporate Data Bases

Forest Service
General Technical Report WO-62



The cover shows an old USDA Forest Service photograph of a range conservationist evaluating one of his sheep. Like a shepherd, the USDA Forest Service has a huge flock of data that must be evaluated before it can be used. This primer provides suggestions for conducting such an evaluation.

United States
Department of
Agriculture

Forest Service

General
Technical
Report WO-62

March 1995

A Primer on Evaluation and Use of Natural Resource Information for Corporate Data Bases

Technical Coordinators

H. Gyde Lund, Research Forester, Forest Inventory, Economics, and Recreation Research Staff, USDA Forest Service, Washington, DC

Charles E. Thomas, Research Forester, Southern Forest Experiment Station, USDA Forest Service, New Orleans, LA

Contributing Authors

Robert G. Bailey, Program Manager, Land Management Planning Staff, Washington Office Detached, USDA Forest Service, Fort Collins, CO

James E. Brickell, Forest Mensurationist (retired), Northern Region, USDA Forest Service, Missoula, MT

William H. Clerke, Remote Sensing Specialist, Southern Region, USDA Forest Service, Atlanta, GA

Robert S. Cunningham, Branch Chief, Programs and Legislation Staff, USDA Forest Service, Washington, DC

Raymond L. Czaplewski, Mathematical Statistician, Rocky Mountain Forest and Range Experiment Station, USDA Forest Service, Fort Collins, CO

Susan J. DeLost, Cartographer, Northeastern Area State and Private Forestry, USDA Forest Service, Morgantown, WV

Fran Evanisko, Cartographer, USDI Bureau of Land Management, Sacramento, CA

Harold R. Greenlee, Cartographer Supervisor, Southern Region, USDA Forest Service, Atlanta, GA

Paul Howard, GIS Specialist, Pacific Northwest Region, USDA Forest Service, Portland, OR

Randy Moore, Deputy Forest Supervisor, Eastern Region, USDA Forest Service, Rolla, MO

Jerry F. Ragus, Regional Soil Scientist, Southern Region, USDA Forest Service, Atlanta, GA

Brian M. Spears, Forest Health Analytical Biologist (GIS), Southern Region, USDA Forest Service, Asheville, NC

Wayne Valentine, Photogrammetrist (retired), Geometronics Service Center, USDA Forest Service, Salt Lake City, UT

Lund, H. Gyde, and Charles E. Thomas (Technical Coordinators). March 1995. A Primer on Evaluation and Use of Natural Resource Information for Corporate Data Bases. USDA Forest Service, Washington Office, General Technical Report WO-62. Washington, DC 168 p., illus.

This primer provides resource inventory specialists with information on how to evaluate existing natural resource information and how to use it in preparing new resource inventories. Subjects covered include determining information needs, finding existing information, determining its utility, evaluating its suitability and quality, and incorporating it into new geographic information systems.

Keywords: corporate data bases, geographic information systems, resource inventories, inventory reports, spatial data bases, nonspatial data bases, maps, overlays, remote sensing, aerial photos, aerial imagery, information sources, data utility, data suitability, data quality, data conversion, data updating

Contents

Preface

The USDA Forest Service is about to embark on an agencywide implementation of geographic information system (GIS) technology and corporate data bases. GIS's and corporate data bases can be invaluable tools for monitoring, protecting, and managing the country's forest and rangeland resources. The value of these tools in meeting specific requirements will largely depend on the quality and usefulness of the information included in the data bases.

Much information exists about natural resources in the form of reports, maps, overlays, imagery, personal knowledge, and data bases. It is usually more economical to use existing information, to the extent practical, than to collect new data, but, in some cases, new data are needed. Evaluation of existing information is important either when designing a GIS or corporate database or when a new information requirement is being defined that will use an existing GIS or corporate database.

What information is needed? What is available? Are the resource data adequate, or do new data need to be collected? These are some of the questions resource and information specialists face. Starting in 1990, the Forest Service's, Washington Office Geographic Information System Steering Committee, through the Timber Management and Information Systems and Technology Staffs, commissioned a task group to develop a primer to address these questions.

During that year, a number of outlines of the proposed primer's content were developed and circulated within the Forest Service for review. Agency specialists in remote sensing, GIS, cartography, information systems, and resource inventory were recruited to write the primer. In May 1993, a draft was sent to selected Forest Service personnel and to internationally recognized authorities for peer review. This volume incorporates the comments and suggestions received from that peer review. It is designed to benefit any person or agency in the process of building corporate data bases, GIS's, or planning inventories.

It is important to learn from the past. When the Forest Service first started the planning process under the National Forest Management Act of 1976, its field units were instructed to use existing information to build their data bases. In some instances, the existing data were outdated or inappropriate for integrated forest planning. Using this data was a costly error resulting in delays in implementation as new inventories were made and forest plans redone. This primer is intended to provide managers and resource specialists with the guidance necessary to build corporate and GIS data bases that will meet the agency's needs now and in the coming century.

Abbreviations

AVHRR	Advanced very high resolution radiometer
BA	Basal area
BW	Black and white (panchromatic) photographs
CFF	Cartographic feature file
CI	Confidence interval
CIR	Color infared
dbh	Diameter at breast height (4.5 ft [1.3 m])
DBMS	Database management system
DEM	Digital elevation model
DLG	Digital line graph
EOSAT	Earth Observation Satellite Corporation
EROS	Earth Resources Observation Satellite Data Center
FIA	Forest Inventory and Analysis Unit
FGDC	Federal Geographic Data Committee
FPM	USDA Forest Service Forest Pest Management
FWS	USDI Fish and Wildlife Service
GIS	Geographic information system
GPS	Global Positioning System
GSC	USDA Forest Service Geometronics Service Center
HRV	High resolution visible
INA	Information needs analysis
IR	Infared
KBS	Knowledge-based system
KSE	Knowledge system environment
MB	Megabyte
MSS	Multispectral Scanner
NA	Not applicable
NAD-27	North American Datum of 1927
NAD-83	North American Datum of 1983
NAPP	National Aerial Photography Program
NFS	National Forest System
NMAS	National Map Accuracy Standards
NRCS	Natural Resources Conservation Service
PBS	Primary base series
PCC	Proportion correctly classified
PLSS	Public Land Survey System
PPS	Probability proportional to size
R-6	USDA Forest Service Pacific Northwest Region
RMSE	Root of the mean square error
SBS	Secondary Base Series
SCS	USDA Soil Conservation Service
SPC	State plane coordinate
SPOT	Système Probatoire D'Observation de la Terre
TM	Landsat Thematic Mapper
USAID	U.S. Agency for International Development
USDA	U.S. Department of Agriculture
USGS	U.S. Geological Survey
UTM	Universal Transverse Mercator

Chapter 1: Using This Primer

Make Believe

This primer provides general guidance on how to locate and evaluate existing natural resource information and how to use such information in the design of new resource inventories. It is intended for resource inventory specialists and for information and resource specialists who seek and evaluate information to enter into corporate (shared) data bases. In this report, we provide boxes highlighting key points addressed in each section. Readers are encouraged to consult the many references provided at the end of this report for more detailed information.

Assume you are the manager of the Enchanted Forest, one of three properties in the Emerald Kingdom of the Imperial Wizard (the Wiz), who is your boss. Other properties include the Deep Dark Woods (a recreation forest) and Sharewood Forest (managed for timber production). You manage the Enchanted Forest for a variety of purposes, including sheep production, timber management, wildlife habitat, and water quality. Until now, all three properties have functioned independently.

Public interest in the administration of the Wiz's properties has increased, and the issues facing her (and consequently you) have become more complex. To better understand what is available and what is happening to the resources in the Emerald Kingdom as a whole, the Wiz wants to create a corporate data base that contains information from all three forests and can be used for a new era of ecosystem management (figure 1).

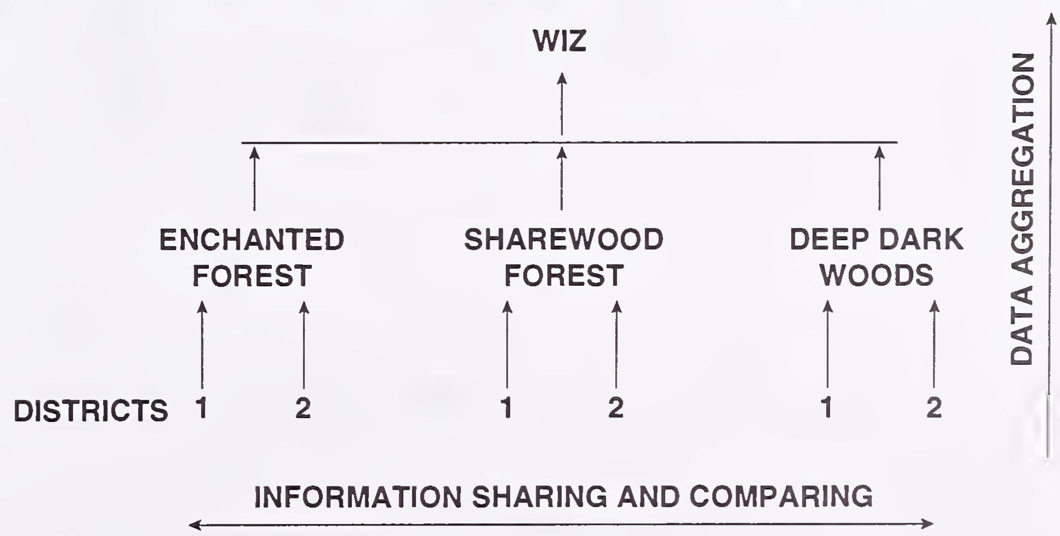


Figure 1—A corporate data base for information sharing, comparing, and aggregating.

At your disposal are hard copies of maps and overlays of various resource themes as well as volumes of inventory reports for the Enchanted Forest. Because your property was managed independently, your definitions and standards may differ from those of the Deep Dark Woods and Sharewood Forest.

Most of your inventories and maps were created 10 to 20 years ago. The accuracy of these information sources when they were created varied from 80 percent to 90 percent. Since their creation, however, changes have taken place. All of the vegetation has changed, of course, due to various processes—grazing, fires, growth, and mortality. About 30 percent of the vegetative cover has been drastically altered due to management activities and wildfire. Some of the changes have been tracked in your data bases through accounting and modeling procedures. But many have not, leaving many of the numerical values in your data bases suspect.

The old data have been the basis for many of your land management decisions. Experience with them has been mixed: sometimes your estimates have been very close to actual results; in other cases the figures have not added up, so to speak. There have been some conflicts with the public over the accuracy of your estimates. Some interest groups have begun producing their own estimates, and they do not always agree with yours.

In addition to your being faced with the task of entering data into the corporate data base, all three forests will be getting Geographic Information Systems (GIS's) (figure 2). Now you must determine whether it is to your advantage to update your existing information, convert it to corporate standards, and enter the maps, overlays, and reports into the GIS, or to collect new information.

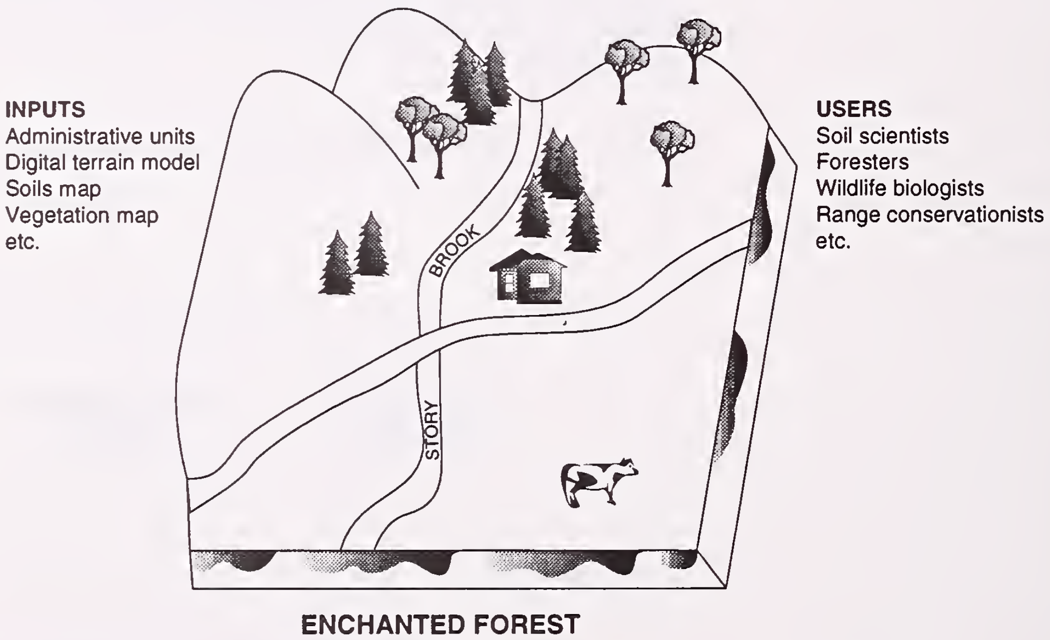


Figure 2—A GIS data base for sharing corporate information among resource specialists.

Some information is available from other sources that may be useful to you. In addition, the technology for obtaining new information has changed drastically in recent years with the advent of digital satellite imagery, videography, and other innovations. Incorporating these new sources of information could enhance your data base and make data entry into the GIS easier.

A corporate data base! Now what?
1) Use old data?
2) Collect new data?
3) Use existing information with new data?
Objectives: low cost, high quality

Like all managers, you have limited funds and personnel, but the wishes of the Wizard must be carried out. What should you do? Convert the old data? Collect new data? If collecting new data, is there any way you can use existing information to help in your inventory efforts?

The purpose of this primer is to help answer these questions. Subjects discussed include determining information needs, consulting sources of information, evaluating data suitability and quality, and using existing information especially for future data collection efforts.

Reality

Management of natural resources has taken on a scope that is unprecedented. Large areas of land and water are manipulated over huge, previously unthinkable spans of time. Whether we realize it or not, we are performing very large-scale experiments in the manipulation of landscapes and ecosystems. An increasingly obvious connection has appeared between management policies and the health or condition of our resource base. The last 50 years have seen tremendous recovery of forest resources in both the northeastern and southern regions of the United States. However, other areas may not have been so fortunate. Resource managers must recognize the uncertainty that accompanies their decisions and begin making management decisions that acknowledge the uncertainties. For too long, we have committed ourselves to looking for the single best management regime—as if we already had perfect knowledge of the complex ecosystems that we manage. This is obviously not the case. Uncertainty is a persistent component of reality. Therefore, managers need to apply a range of management options for lands and water and to treat these applications as learning opportunities by carefully observing their effects. Application of one alternative isn't necessarily a last opportunity—it can still be monitored and learned from—but it does reduce learning opportunities. The practice of applying only a single management regime over large areas needs to be reexamined and a new multihypothesis management tried, along with an improved scientific approach to recording and observing the outcomes of alternative management regimes (Walters and Holling 1990). Knowing the current resource condition and monitoring its development are crucial to this new approach to managing our natural resource base.

Agencies with large land bases and complex information needs are developing corporate data base systems designed to eliminate the following problems (modified from Aronoff 1989):

- Data that are poorly maintained or out of date
- Data that are not recorded or stored in a standardized way

- Data that are not defined in a consistent manner

- Data that cannot be shared, compared, or aggregated

- Systems that have limited data retrieval and manipulation capabilities, and

- Systems that cannot meet the new demands that are being placed on organizations

The USDA Forest Service, like many other public agencies, has entered the standardized corporate data base and GIS era. The standardized data base will be essential in sharing information and experiences throughout the agency. The GIS will be an invaluable tool for helping decisionmakers to manage and protect the Nation's natural resources (figure 2). However, the results from a GIS and the supporting corporate data bases will be no better than the data entered.

The value of existing information for future use needs to be considered by both decisionmakers and practitioners. In some circles, earlier efforts or other agencies' projects are often criticized or even ridiculed. But current scientific and statistical methods are making it increasingly easy and appropriate to incorporate past information into current efforts. Cultivating an attitude of respect for existing information will provide previous efforts with the acknowledgment they deserve while improving the products needed to do our job now.

Much information exists about natural resources in the form of reports, maps, overlays, imagery, personal knowledge, and data bases. Several factors are making the quality of resource information more important than in the past.

Factors influencing Forest Service inventory processes:

- Reduced budgets demand more effective use of our resources for data collection and processing.

- The need to integrate resource information across space and time requires compatible data sets and consistent application of evaluation criteria.

- New decisionmaking strategies (i.e., ecosystem management) rely on geographically referenced information to achieve an acceptable blend of resource uses.

- More sophisticated information-processing technology places our basic data and decisionmaking assumptions under greater public scrutiny.

Managing public lands is increasingly complex. To address the complex problems, the Forest Service and other agencies have amassed large amounts of data and information related to natural resources. It is urgent that agencies base land management decisions on the best information available. It is also important for agencies to have sound methodologies for testing data sets in relation to their intended use. A summary of such methods appears in the following box.

Methods of reviewing data adequacy for inclusion in corporate data bases:
1. Identify problems and inadequacies before automation. After data are automated, problems are difficult and expensive to fix.

2. Assess the distribution and causes of errors to provide a basis for developing better data bases in the future. Detection of systemic errors, for example, may provide clues to processing problems or point to variables or associations previously unknown.

3. Understand the nature of errors to provide a basis for assessing the uncertainty associated with operations of the data base. Knowledge of uncertainty reduces inappropriate data use and helps in choosing among alternative approaches to problem solving.

Thus far, the Forest Service has focused on information system architecture and developing standard terminology for describing ecological variables (USDA Forest Service 1988a and b). Recently, several Forest Service authors (Bailey 1988b, Lund 1986a and 1990, Evanisko 1990) have stressed the importance of the quality of natural resource data bases, especially those intended for use in an electronic setting for corporate GIS's.

Bailey (1988a) points out that representing ecological units as uniform regions may lead to false conclusions. Such representations may not capture significant subunits of productivity or ecological response. He suggests placing "less attention on the technology and more on getting better information." Unfortunately, getting better information is not simple. Data quality can only be judged in terms of specific operational goals, such as improving wildlife habitat or increasing the quality of water. The scale of analysis and the local geographic context in turn influence objectives. For example, are we trying to improve wildlife habitat in a particular watershed, national forest, or State? Each level could have different data needs. Rigidly uniform accuracy standards that ignore scale and geographic context will not work for data bases depicting ecological variability.

Recently, there have been requests to examine existing information to determine its utility in corporate data bases and GIS's (Lund 1990, Winterberger and Reutebuch 1990). We prepared this report to meet those requests.

This primer provides guidance on how to:
- Determine information needs
- Locate existing information
- Evaluate existing information for use in corporate and GIS data bases, and
- Use existing information in new data collection efforts.

This is the third of a series of primers dealing with resource inventories. The first (Lund 1986a) dealt with integration of inventories, and the second (Lund and Thomas 1989) addressed a variety of inventory designs in use by the Forest Service. This primer reviews information needs assessment techniques, evaluation of existing information, and use of information in corporate data bases, including those needed

for GIS's. This report also provides general guidance to those who face the difficult decision of whether to use existing data or to go out and collect new information. Lastly, this primer provides guidance on how to use existing information in new data collection efforts. A glossary is provided at the end of the document defining cartographic, remote sensing, resource inventory, and data management terms used in this report. Also included are numerous references to works the reader may consult for more detailed information.

Chapter 2: Determining Information Needs

As manager of the Enchanted Forest, you generally need to know (1) how much of a resource there is, its condition, and its location, (2) what the potential of the land and resource base is under various management alternatives, and (3) what the suitability is of the land and resource for management. The exact information you need depends on what decisions are to be made and how the data are to be used.

Examples of information needs for:
Inventories: Census of discrete objects; estimates of discrete objects; estimates of continuous univariate distributions; estimates of multivariate distributions.

Evaluations of potentials: Landslide potential; erosion hazard; regeneration potential; natural vegetation potential.

Evaluations of suitability: Wildlife habitat suitability; suitability for specified management activities and practices.

Needs for Resource Management Decisions

Whether you are manager of the Enchanted Forest or Imperial Wizard of the Kingdom, you have to know what information you will need to make decisions. To determine information needs, you must (1) identify the questions to be asked and the management decisions to be reached; (2) identify and characterize the data needed to reach those decisions; and then (3) select the right information to gather. This process is called an information needs analysis or assessment (INA).

Hoekstra (1982) and Lund (1985, 1986a, 1987) provide detailed instructions for determining information needs for large public agencies and industrial organizations where corporate data bases are essential for overall planning and reporting. The steps include (1) reviewing laws, regulations, cooperative agreements, and memoranda of understanding to identify the information required at the broadest level of the organization; (2) examining emerging issues both nationally and locally; and (3) looking at data the decisionmaker needs in order to manage the resource at the local level.

Within the Forest Service, the National Forest System (NFS) currently consists of the Washington Office (national headquarters), 9 regional offices, 155 national forests and 20 national grasslands, and more than 600 ranger districts. Managers need data at all levels for national planning and reporting to the district for operations. Data collected at any level, including district data collected for upward reporting, are corporate information.

In an agency like the Forest Service, laws and directives spell out some of the corporate information needs. Thus, laws and directives are among the first things to check for required information. Additional needs accumulate as one moves from the highest echelons in the agency to the lowest. The following box contains key questions to ask in determining information needs.

Questions to ask before assembling a new inventory:

- What do laws, charters, or higher echelons of the organization require, and what data are needed to meet those requirements?
- What current and future issues and resource decisions does the manager face, and what additional data are needed to face them?
- What is the geographic area in question?
- What is the risk (cost) of an incorrect decision? How accurate must the data be?

To answer these questions, the resource specialist or person responsible for obtaining data must understand the decisionmaking process, identify who makes decisions, identify other parties involved, and involve these individuals in defining information needs.

The next step is to identify information that is needed to address issues and solve problems. First the needed models, tables, maps, data bases, and report forms must be developed. Then data elements needed to generate the required information must be identified. This step should be done as precisely as possible. For example, if a decisionmaker needed to know the area of fictitious spotted-snail eater habitat in the Enchanted Forest, key habitat elements (such as vegetative cover and size of area) would have to be defined to determine information needs.

Before progressing further, the decisionmaker should perform an INA, reviewing and approving the required tables, maps, data bases, and data elements. An example of such an INA is provided in USDA Forest Service (1990b), supplemented by common standards and definitions for use throughout the organization (USDA Forest Service 1989 and 1990c).

The INA is not a one-time activity. As programs and activities change, the INA process must be repeated to identify additional or modified information requirements.

In the initial stages of gathering data or developing an inventory process, the natural resource issues that may emerge must be carefully considered. Because there is no such thing as perfect foresight, we make our best estimate of the issues that will confront us in the future. In addition, we need to consider how decisionmakers will resolve these issues. Projections of future issues and their resolution can be enhanced through effective teamwork and interaction with natural resource interests as well as thoughtful consideration of our past land management problems. This allows resource teams to reach conclusions on the necessary parameters of inventory and information development. Teams should be able to define the types of data they will evaluate, the needed geographic area of concern, and the required precision and reliability of selected parameters.

The level and reliability of information require very careful consideration and evaluation before significant resources are committed to gathering data. The success of many projects—in a technical as well as a social context—rests upon effective

information management, which requires close communication and interaction with those involved in the administration of natural resources. Interaction must be based on the recognition that development and use of information are social processes aided through mutual dialogue and understanding. Potential users must keep in mind that all necessary information is seldom available, and that available information might not be as accurate as one might wish. Potential liabilities vary with the amount of usable information and its relative accuracy.

In the early stages of data collection and interpretation, the relative risk of an incorrect decision must be weighed against the cost of information in dollars, time, and personnel. After practitioners and managers decide what issues to expect and how to resolve them, the consequences of using incomplete or inaccurate information must be determined and evaluated. Inadequate information could damage relationships with the public.

Needs Specific to Corporate Data Bases

Corporate data bases include data from various components of the agency for sharing, comparing, and aggregating. Data such as vegetation type and topography that are common to various resource sectors may be shared by the wildlife specialist, soil scientist, and forester. These data are one form of corporate information.

Data entered into a data base for comparison and upward reporting are another form of corporate data. The same or similar data are collected in parallel units of the organization (i.e., districts, national forests, and regions). Administrators use corporate data bases primarily for comparing resources in parallel administrative units and for aggregated reporting. Complete data sets following standard definitions and coding are essential to effective data base functioning.

The Forest Service is attempting to develop an information base that responds to a well-defined core set of questions. For this to work, our corporate data must be aggregable across space, through time, between resources, and at various levels within the agency (Lund 1987). Aronoff (1989) lists the following advantages and disadvantages of data bases that have been modified to apply to corporate needs:

Advantages of Corporate Data Bases

- Centralized control, which ensures that data quality standards are maintained, security restrictions are enforced, conflicting requirements are balanced, and data base integrity is maintained

- Flexibility, which fosters the development of new applications through data handling services

- Independence of application programs from the physical form in which data are stored

- Easy implementation of new application programs and unique data base searches, and

- Elimination of redundancy.

Disadvantages of Corporate Data Bases

- Cost of data base system software and any associated hardware needed (at a minimum, existing facilities may require upgrading and increased maintenance costs)

- Increased susceptibility to failure and difficulty of recovering data lost due to complexity of corporate data base system, and

- Risk of loss or corruption of data due to centralizing their location and reducing their redundancy.

The second and third disadvantages may be minimized by effective backup and recovery systems (Aronoff 1989).

Needs Specific to Geographic Information Systems

Data stored in a GIS are a special form of corporate data: multiple themes are stored spatially in a common data base for sharing among a variety of users. Geographic Information Systems are well suited to natural resource management, providing the information and analysis tools necessary to support activities common to all natural resource management organizations. The most basic function of a GIS is to provide a description of current environments. A GIS can also be used to identify management direction and to develop plans. Information about the existing environment is essential but not necessarily sufficient for developing plans. Additional data may have to be acquired. A GIS can also assist the manager in implementing management plans and monitoring the effectiveness of management activities.

Basic GIS characteristics for management use:
Geographic Information Systems are used to store, display, and analyze the spatial relationships of features in reference to the Earth's surface. Two kinds of data are stored: spatial coordinates describing each feature's location; and attributes describing its characteristics.

Geographic Information Systems store feature locations in either a vector or raster format, although recent developments have led to hybrid systems. In vector systems, the shapes of features are stored in coordinate pairs. Point features (such as wells or plot centers) are identified by a single coordinate pair; linear features (such as roads or streams) are identified by a string of coordinate pairs; and polygon features (such as stands or lakes) are identified by the coordinate strings that delineate the polygon boundary. The coordinates of each feature are explicitly defined and stored in the system, either as geographic coordinates (of latitude and longitude) or as plane coordinate referenced to a map projection.

In raster systems, features are identified by their location in a rectangular data structure composed of rows and columns of data values. Point features in a raster system by a single cell, linear features by a string of adjacent cells, and polygon features by a cluster of adjacent cells (Aronoff 1989) (see figure 3).

THE RASTER AND VECTOR DATA MODELS

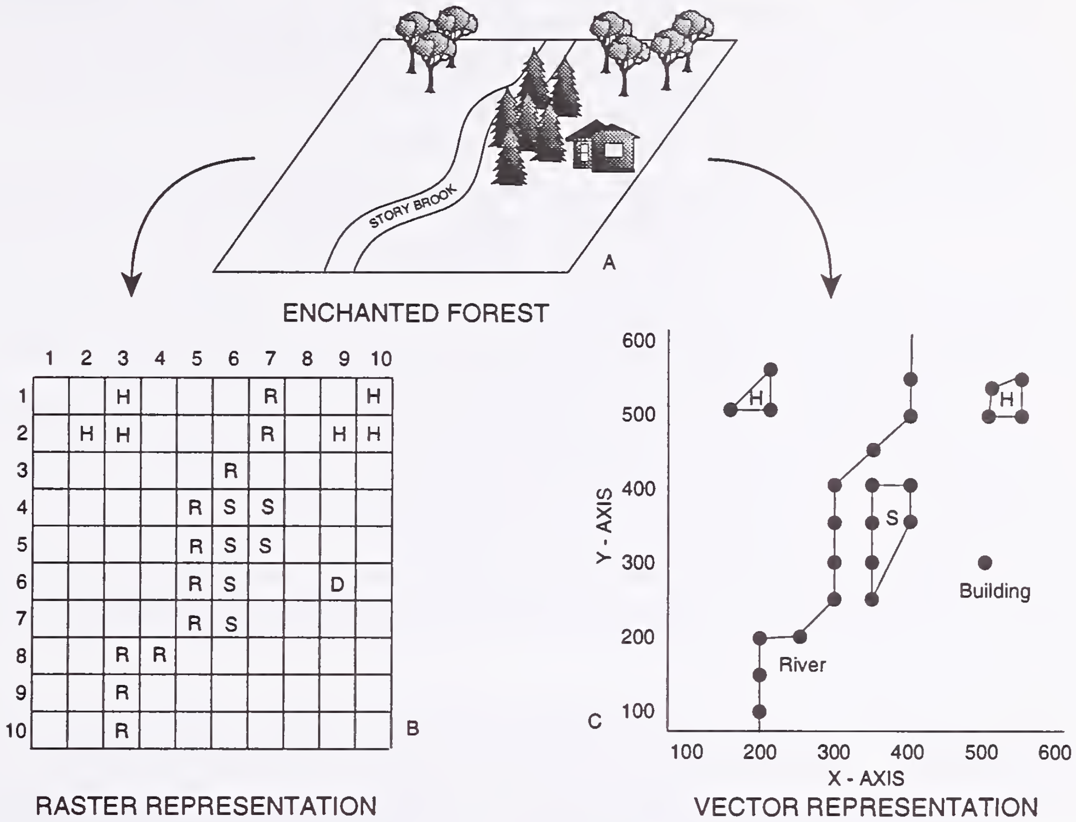


Figure 3—Comparison of vector and raster GIS models (modified from Aronoff 1989). A portion of the Enchanted Forest (**A**) is shown in raster representation (**B**) and in vector representation (**C**). The softwood (S) and hardwood (H) stands are area features. The Story Brook River (R) is a line feature, and the building representing district headquarters (D) is a point feature.

Geographic features in a GIS are linked to one or more descriptive characteristics of the feature. In a raster GIS, the numeric value stored for a raster location may identify the class or characteristic of the feature or serve as a link to descriptive information stored separately. In a vector GIS, a link stored with the coordinate data provides a pointer to one or more characteristics of the feature stored within the GIS or an associated data base management system (DBMS).

Geographic Information Systems features must be referenced to a location on the surface of the Earth. In a timber inventory, for example, tree measurements are often made at points clustered around the plot center. All the measurements on the plot are generally referenced to the coordinates of the plot center. Depending on analysis requirements, however, the tree measurements could be referenced at each cluster plot to the coordinates of the cluster center. For more specialized stand development studies, it would be possible to relate the measurements of individual trees to their geographic location. At the other end of the scale, summary inventory statistics for an administrative or political unit can be stored in a GIS referenced to the coordinates defining the unit. In these cases, a standard code, such as the Federal Information Processing Standards county code, could serve as the link between many individual attribute data sets and a geographic theme of county boundaries. Geographic Information Systems data differ from those of other kinds of data bases in that the information must be tied to specific locations on the ground through various coordinate systems.

What information should the GIS include? One approach would be to include all information available for the area of interest (national forest or ranger district). This approach would increase the cost of data preparation and storage. The appropriate strategy is to include in the GIS those data that are necessary to support the users' information needs.

The INA is a widely accepted approach for identifying data that should be included in a GIS. The objective of the INA is to identify the individual data sets that should be included (such as streams or transportation), but also the characteristics and reliability of the data. The INA for a GIS includes both the spatial data (points, lines, and polygons) that define spatial location, and the attribute or tabular data that describe these spatial locations. This is done by taking a product approach. Participants are asked to brainstorm to come up with the major management concerns and issues that pertain to the unit (e.g., district project area). They identify the products (such as maps, tabular reports, and graphs) that are needed to address the concerns or issues. These are ordered by priority, and the specific data (both spatial and attribute) needed to make the product are identified (Anonymous 1991).

At the national level, a GIS is likely to consist primarily of aggregated attribute data and generalized geographic representations summarized from more detailed information used at the operating unit level within the organization. In order for information from GIS's to be aggregated across administrative units, basic data elements should have common definitions. Functionally compatible models should be used to develop the information to be included in the corporate data base. At the operating unit level (district), individual features recognizable on the ground are stored in the GIS to support the requirements of tactical planning and analysis (see Valentine [1990] and Evanisko [1990]).

Geographic Information Systems operated by natural resource agencies generally include three data types: base, resource, and derived. Base data include the thematic data layers or themes describing the basic characteristics of the unit. These include transportation, hydrologic, and administrative boundary themes. Medium-scale planimetric or topographic maps are a good source of base data. For example, the Forest Service uses a modified version of the USDI Geological Survey (USGS) 1:24,000-scale, 7.5-minute quadrangle as the basis for field activities, except in Alaska. To create Forest Service primary base maps, information is added to the USGS quadrangles to depict and identify the transportation network, ownership, administrative sites, and administrative boundaries. Mylar stable base composites of these maps are digitized using tablet digitizers to create digital cartographic feature files (CFF's). These data sets include the point, linear (except for contours), and polygon features for symbols on the maps, along with location information.

The CFF serves as an interchange file providing a complete data set in a simple format that can be restructured to the specific requirements of a digital cartographic system or GIS. Each feature is tagged with one or more cartographic codes identifying the features associated with the coordinate data. A stream might have a single feature code, whereas a river that was also the boundary of a county would have two codes—one for each feature represented by the coordinate string. Elements are extracted from the CFF by feature code to create individual GIS themes.

The CFF will soon be completed for the national forests. These files comprise the cartographic layers of individual 7.5-minute quads that have been updated by the Forest Service. The layers are: hydrology, transportation, Public Land Survey System (PLSS), boundaries, and culture. Point features in these data files generally are attributed completely, but for cartographic purposes only (e.g., symbology and orientation). Linear features are somewhat attributed, but where the linear feature is also connected with a polygon (e.g., double-banked streams), essential topology is lacking. No directional sense is coded for individual vectors.

Polygon feature attributing is lacking. For example, there is no topology associated with PLSS or boundary sets, nor are multiple codes used (e.g., a road on a section line that is also a county boundary). Therefore, when translating from a CFF to a digital line graph (DLG), nothing may get translated. An element in the DLG is an empty set if there is no topology in the CFF. While correct insofar as location is concerned, these files are limited for GIS use because their attributing is limited and topology is nonexistent. Full GIS utilization of these data will require varying degrees of additional attributing and complete topology. GIS users need to be aware of this.

Timber stands and soils are examples of resource themes. Resource data are drafted on overlays registered to the Mylar primary base map. Boundaries coincident with a base theme are drafted separately from noncoincident portions of the resource overlay. To create the timber stands theme, the noncoincident portions of the stand boundaries are digitized, and the coincident portions of the boundaries are copied up from the base layers. This approach to GIS data base construction provides a vertically integrated GIS in a single coordinate string representing a feature on all pertinent themes.

After each theme has been constructed, it must be labeled and linked to the appropriate attribute data. The usefulness of a GIS data base depends on both the accuracy and the timeliness of the data themes. Updating the GIS to represent current conditions is a continual process. Updates may be added as events occur or on a cyclical basis. GIS data bases that are not maintained soon lose their usefulness.

Data appropriate for the level of planning and analysis are needed. Primary base series (PBS) data will meet the needs for strategic levels of planning, such as developing forest plans. However, a GIS is a powerful tool that can be used for project planning. Table 1, for example, lists soils information needs for a variety of uses. Note that the minimum size shown is about the smallest delineation allowable for readable soil maps. In practice, the minimum size delineations are generally larger than the minimum size shown. The level of data resolution required for "project planning" varies with the scope of the project.

An example of project planning on a strategic level is a prescribed burn for controlling western juniper. Islands of ponderosa pine and a powerline on wooden poles occur within the project perimeter. All buildings are outside the perimeter, but a few are nearby. Themes needed in the GIS to plan this project are vegetation types, fuel loading, contour lines, fences, roads and trails, pipe and transmission lines, streams and water sources, historical and cultural sites, threatened and endangered species,

Table 1—Levels of soils inventory within the Forest Service.

Order	Data needed	Field procedure	Minimum area (ha)
I	Very intensive (e.g., experimental plots, individual building sites)	The soils in each delineation are identified by transect or traverse. Soil boundaries are observed throughout their length. Remotely sensed data are used to assist boundary identification.	1 or less
II	Intensive (e.g., general agriculture, urban planning)	Same as for Order I. Soil boundaries are verified at close intervals.	0.6–4
III	Extensive (e.g., rangeland, forest land, community planning)	The soils are identified by transect of representative areas, with some additional observation. Boundaries are plotted mostly by interpretation of remotely sensed data and verified by some observations.	1.6–256
IV	Extensive (e.g., regional planning)	The soils are identified by transect of representative areas to determine soil patterns and composition of map units. Boundaries are plotted by interpretation of remotely sensed data.	40–4,000
V	Very extensive (e.g., selections of areas from more intensive orders)	The soil patterns and composition of mapping units are determined by mapping representative areas and applying the information to the areas by interpretation of remotely sensed data. Soils are verified by occasional on-site visits or traverses.	1,000–4,000

and buildings. Slope and aspect are derived from contours. This information is used with data on fuel loading and vegetation type in formulation of the burn prescription. Vegetation types, fences, pipe and transmission lines, streams and water sources, historical and cultural sites, threatened and endangered species, and buildings show us the location and status of entities that need to be protected. Roads and trails indicate existing firelines. We can then decide where new ones are needed. Streams and water sources indicate where water is available for fire control.

Planning is very complicated, with all the entities requiring protection by law. The GIS is valuable because of the speed with which all of these entities can be overlaid and viewed simultaneously, plans formulated and drawn, and changes made. Functions such as these may take months to do manually or be impossible if too complex. They can be completed in a matter of hours with a GIS, provided the data are available, current, in the GIS, and of sufficient quality and resolution.

A GIS can be used to expedite land management planning, legal proceedings, and the writing and illustrating of environmental documents. Spears and Nettleton (1993) reported using a GIS to produce graphic products involved in managing a southern pine beetle outbreak that was threatening red-cockaded woodpecker colonies in and adjacent to the Little Lake Creek Wilderness in Texas. The graphics were used in litigation.

Needs Specific to Future Resource Inventories

Existing information is essential for planning new inventories. Information about the location and changes in the resource base, roads and trails, etc., can help the inventory specialist. Personal knowledge of an area can help focus activities on areas of known change and help plan logistical aspects of data collection. Past inventory documentation and plans can provide links for monitoring and change detection. Existing maps, overlays, and remote sensing can be used to stratify the land and resources, making data collection efforts more cost-effective.

Summary

Natural resource inventory data in the form of maps are often the first point of contact in making choices about natural resources. Users must understand the information at hand before collecting more data. Data interpretations by professional scientists and practitioners are critical in facilitating comprehension of the value of inventory information. Mutual understanding of the management decision contemplated is crucial for the efficient and effective use of information. Information needs assessments must include considerations of costs, issues to be resolved, and risks of incorrect conclusions. Integrated teams of managers, practitioners, and interested citizens should make the choices for the collection and interpretation of information. Development and use of information are social processes aided by mutual dialogue and understanding.

Chapter 3: Finding Sources of Information

As discussed in the previous chapter, the INA provides the basis for undertaking a search for existing data.

Kinds of Information

The manager of the Enchanted Forest will find that existing information is abundant, but often diverse and scattered. It may take the form of personal knowledge, inventory reports and data bases, maps and overlays, remote sensing products, and other georeferenced data.

Forms of existing information include:

- Personal knowledge
- Inventory reports and data bases
- Maps and overlays
- Computer spatial data bases
- Remote sensing products

Personal Knowledge

Probably the most often-used sources of information are personal contacts and knowledge. These sources are also probably the least documented and most difficult to evaluate.

Personal knowledge is frequently used to locate and evaluate data sources. Many sources of data, particularly those developed as part of specialized or academic studies, are not included in published listings of available data. Individuals working in the discipline or geographic region of interest are frequently aware of unpublished data sources. Personal knowledge is also important in identifying the lineage and characteristics of data sets. This is especially true of nonstandard data sources that may not be well documented. "Old timers" in field units are good sources of historical information and may be able to identify additional data sources.

Personal knowledge may be the only or most readily available source of information on past conditions or the occurrence of rare phenomena. In cases where no other data source is available, it may be possible to develop a data layer from an individual's memory of past conditions. For example, the U.S. Agency for International Development (USAID) needed information on past vegetation conditions in a part of Sudan for rehabilitation work. The only source of information was the farmers and villagers who lived and worked in the area. Through personal interviews, USAID got the needed estimates of past vegetation conditions and rates of change (Lund and others 1990).

With current technology, it is possible to incorporate personal or qualitative (heuristic) knowledge into an analysis system along with conventional sources of information such as inventory reports, maps, and satellite imagery. A knowledge-based system (KBS) uses expertise and heuristic knowledge as a basis for analysis, just as a GIS uses spatial information. The two major components of a KBS are the knowledge base and the inference engine.

Knowledge engineering is the process for assembling and organizing information into a knowledge base for specific problem solving. Production rule systems are a commonly used method of structuring qualitative information in a knowledge base. According to Chen and others (1991), "A production rule system consists of a currently perceived state or context ("if"-component), the goals of the individual, an appropriate action ("then"-component), and a state the decisionmaker expects to reach if the action is taken."

The user interacts with the knowledge base through the inference engine to obtain a result or recommended action. For example, the user might wish to know the appropriate thinning density for a plantation. When only a single stand is being considered, the system might request specific information about the stand. Depending on the answers provided by the user, additional rules would be triggered and additional information requested from the user. Where many stands must be evaluated, it is more efficient to link the KBS with a DBMS containing the relevant characteristics of the stands. In this case, the user could identify the stand explicitly (stand number) or implicitly by condition or age class. In many cases, it is inefficient or impossible to store all the relevant information needed to reach a decision about a feature of interest, such as attributes of a timber stand, in a tabular DBMS. This is true, for example, when adjacency must be considered. In these cases, the KBS must be able to access a GIS describing the environment as well as the attributes of the stand stored in the DBMS. Proximity to a stream or the type of the soil on which the stand is located are environmental variables that might affect the suggested thinning regime.

The integration of DBMS, GIS, functional models, and a KBS is described as a knowledge system environment (KSE). The Forest Service is developing several decision systems based on this technology. The Integrated Southern Pine Beetle Expert System (Texas A&M University 1992) and the Integrated Forest Resource Management System–Texas (USDA Forest Service and others 1992) are examples of KSE technology being developed by the Forest Service.

Inventory Reports and Data Bases

The next most common source of existing information comes from inventory reports and data bases. For example, the Forest Service's Forest Inventory and Analysis Units (FIA's) have regularly produced reports for the Eastern United States since the 1950's (and irregularly before that time). Rosson and others (1988) provide timber inventory information for Louisiana that is typical of the kinds of published information available from the FIA's. Consumers, however, should be aware that the publication date (1988) is different from the resource evaluation date, in this case, 1984. Thus, the age of the information is often important in deciding whether or not to collect new primary or auxiliary data. Good inventory reports should also contain the statistical errors associated with the combined statistics. For disaggregated data, errors may or may not be available according to the number of plots that may occur in a given mapped polygon or area.

Plot-level information, used to assimilate reports, may also be available in a variety of electronic transfer media such as tapes and diskettes. With the advent of computers and various data base management tools, FIA data have gained some timeliness.

In addition, data from a plot-level or tree-level data base may be available to those who assemble a GIS. The extremely low intensity of FIA sampling over the inventory unit means that the number of plots will probably not sample all the conditions on national forests, though there may be closely located plots that could be substituted. These kinds of data bases are complex. Potential users should seek the help of the people knowledgeable about the data base before trying to use them. Often, terminology may mislead potential users if not thoroughly understood.

Maps, Overlays, and Computer Spatial Data Bases

Maps and overlays are also quite common. Spatial information is essential for many resource analysis tasks. Traditionally, in cartography, there are “general purpose” and “special purpose” maps. General purpose (reference) maps convey knowledge about the location of geographic features such as topography and transportation nets. Special purpose (thematic) maps or overlays, on the other hand, convey one idea well.

Most maps and overlays will eventually be converted to a digital format. Historical maps are less likely to be available in digital form. Today, it is difficult to make a clear distinction between maps and digital representations of the Earth’s features. It is often impossible, looking at the final product, to determine if a map was produced using manual or digital cartography. Manually generated maps can be converted to digital data to support later cartographic and GIS application. The Forest Service is now digitizing data on all primary base quadrangles within the NFS. The USGS has a goal of converting the 1:24,000-scale quadrangles nationwide to a digital format by the year 2000. Data files to support cartographic or GIS application differ in structure and format rather than content.

Within the Forest Service, the main sources of data for resource planning for a GIS are PBS maps, along with their resource overlays. Primary base series maps represent the best source of digital geographic data available for a GIS in a reasonable time frame. Moreover, PBS data are probably more consistent and accurate than resource data. These data apply to such activities as strategic forest planning and cumulative effects analysis, where practical limits to map accuracy and content are of little or no consequence.

Information on various themes are collected from resource inventories, remotely sensed data, and mapping activities. These data are commonly overlaid in a GIS to derive new information. Compared to other corporate information, data precision, resolution, and quality are more important for spatial data bases than standardization across functions (Valentine 1990, Evanisko 1990). If precision, resolution, and quality of the overlaid themes differ, erroneous interpretations result.

Remote Sensing

Remote sensing data (including historical photographs) are essential for mapping and natural resource management. Remote sensing systems, commonly used to support resource management activities, sense electromagnetic energy emitted or reflected from objects in the environment. Aerial photographic and video systems sense reflected energy in the visible and near-infrared (IR) portions of the spectrum. Electro-optical sensors carried aboard aircraft and satellite platforms sense reflected

energy from the visible blue to the thermal IR portions of the spectrum. Airborne and satellite radar systems are active systems sending out pulses of microwave energy. Radar data are based on the characteristics of the energy reflected from scene elements. Because radar is an active sensor, it is capable of acquiring imagery through cloud cover or in darkness.

It would be virtually impossible to assemble a GIS without data derived from remote sensing systems. Aerial photography, scanners, and video systems are widely used to acquire spatially referenced data for preparing the base cartographic data, for mapping, and for inventorying natural resources. No single sensor system is suitable for all applications. Panchromatic medium-scale aerial photography is essential for base map construction and update. It is also used to create orthophotos commonly used by resource managers for vegetation delineation. Medium- and large-scale color and color infrared (CIR) aerial photography are widely used in delineation of resource features such as timber stands and soils. Cover class delineation over extensive areas and change detection can be readily performed using automated procedures and digital satellite imagery. Video systems can be mounted in small aircraft to acquire imagery with no delay for processing.

Acquiring and extracting information from remote sensing imagery requires significantly more effort and expense than automating an existing map, which is basically a task of data reformatting. On a map, the information of interest has already been extracted from the source material, categorized, registered to geographic coordinates, and, in most cases, verified. To extract information from remote sensing imagery, the user must perform all these tasks and often be responsible for mission planning and image acquisition. The cost and effort required to extract information from remote sensing imagery should be carefully weighed against the quality, content, and timeliness of existing data. Elevation data, for example, are available for many areas in digital data sets at several resolutions and extents of coverage. If these data do not provide sufficient detail for a specific application, new data can be acquired from stereo aerial photography or satellite imagery. Similarly, when existing data sets or maps do not include sufficiently detailed or updated data on cultural features, it may be necessary to extract the information directly from remote sensing imagery.

Table 2a reviews and compares some of the most commonly used remotely sensed data types, both digital and photographic. More discussion of various sensors and platforms may be found in the Appendix.

Table 2a is only a brief guide to these data sources, and further research should be done before considering any source for a particular application. Table 2b provides an overview of different remotely sensed data types and their potential contribution to a GIS base layer creation.

Some of the themes or layers presented (such as vegetation height) require high resolution data to make precise quantitative measurements, whereas others (such as snow cover) do not. Application requirements should be determined before evaluating available sensors and data.

Table 2a—Characteristics of commonly used remotely sensed data types based upon Lachowski (1990).

Characteristics	AVHRR [a]	Landsat MSS [b]	Landsat TM [c]	SPOT/ms [d]	SPOT/pan [e]	NAPP [f] 1:40,000	Resource photography [g] 1:24,000	Resource photography [g] 1:12,000	Video [g]
Spatial resolution	1.1 km	80 m	30 m	20 m	10 m	20 lines per mm	NA	NA	NA
Spectral range	.58–1.1 (2 bands)	.50–1.1 (4 bands)	.45–12.5 (7 bands)	.49–.91 (3 bands)	.51–.73 (1 band)	.4–.9	.4–.9	.4–.9	.4–.9
Type of composites possible	Color/BW Color-IR	Color/BW Color-IR	Color/BW Color-IR	Color/BW Color-IR	BW-IR Color-IR	Color/BW Color-IR	Color/BW Color-IR	Color/BW Color-IR	Color Color-IR
Area of coverage (km per scene/ photo)	2,400 x 2,400	185 x 170	185 x 185	60 x 60	60 x 60	9.1 x 9.1	5.5 x 5.5	2.6 x 2.6	Variable
Available as digital (without additional processing)?	Yes	Yes	Yes	Yes	Yes	No	No	No	No
Cost of scene as digital	$25	$1,000	$4,900	$2,200	$2,400	NA	NA	NA	NA
Cost per 7.5' quad for digital or number of photos	$0.39	$4.45	$40.62	$69.55	$98.55	10 photos	20 photos	71 photos	NA
Cost of hard copy per scene/photo	NA	$1,000 (1:250,000)	$1,500 (1:250,000)	$1,400 (1:200,000)	$1,500 (1:200,000)	$8	$4–8 (Color-IR)	$4–8 (Color-IR)	NA
Frequency of coverage	Daily	16 days	16 days	26 days	26 days	5 yrs	Variable	Variable	Variable
Typical hardware for interpretation	Computer/ software	Computer/ software	Computer/ software	Computer/ software	Computer/ software	Stereoscope & light table	Stereoscope & light table	Stereoscope & light table	VCR & computer software, hardware
Digitizing required	No	No	No	No	No	Yes	Yes	Yes	No

a. Advanced Very High Resolution Radiometer, available through EROS Data Center.
b. Landsat Multispectral Scanner (4 bands), available from EOSAT and EROS Data Center.
c. Landsat Thematic Mapper (7 bands), available from EOSAT and EROS Data Center.
d. SPOT multispectral (3 bands), available from SPOT Image Corporation.
e. SPOT panchromatic (1 band), available from SPOT Image Corporation.
f. National Aerial Photography Program, available from Aerial Photography Field Office (APFO).
g. Resource photography and video acquisition must be coordinated by forests.

NA Not applicable.

Note: The products presented are typical data formats that have been used for creation of GIS databases.

Table 2b—Recommended uses for remotely sensed data spirces and GIS data base creation based on Lachowski (1990).

Features	AVHRR	Landsat MSS	Landsat TM	SPOT/ms	SPOT/pan	NAPP	Resource photography 1:24,000	Resource photography 1:12,000	Video
*Natural**:									
Basal area	0	3, b, c	2, 3, b, c	2, 3, b, c	0	2, 3, b	2, 3, b	1, 2, b	1, 2, b
Canopy cover	3, b	2, 3, b, c	1, 2, 3, b, c	1, 2, 3, b, c	0	1, 2, 3, b	1, 2, 3, b	1, 2, b	1, 2, b
DBH (size class)	0	2, 3, b, c	1, 2, 3, b, c	1, 2, 3, b, c	0	1, 2, 3, b	1, 2, 3, b	1, 2, b	1, 2, b
Species	0	3, a, b, c	2, 3, a, b, c	2, 3, a, b, c	0	2, 3, b	1, 2, 3, b	1, 2, b	1, 2, b
Existing vegetation	3, b	2, 3, a, b, c	1, 2, 3, a, b, c	1, 2, 3, a, b, c	0	1, 2, 3, b	1, 2, 3, b	1, 2, b	1, 2, b
Vegetation height	0	0	0	0	0	2, 3	1, 2, 3	1, 2	1, 2, b
Vegetation density	3, b	2, 3, c	1, 2, 3, c	1, 2, 3, c	1, 2, 3, c	2, 3	1, 2, 3	1, 2	1, 2
Snag condition	0	0	0	0	0	2, 3	2, 3	1, 2	1, 2
Forest/Nonforest	3, b	1, 2, 3, c	1, 2, 3, c	1, 2, 3, c	1, 2, 3, c	1, 2, 3	1, 2, 3	1, 2	1, 2
Hardwood/conifer	3, b	1, 2, 3, b, c	1, 2, 3, c	1, 2, 3, c	1, 2, 3, c	1, 2, 3	1, 2, 3	1, 2	1, 2
Structure (forest)	0	0	2, 3, b, c	2, 3, b, c	1, 2, 3, b, c	2, 3	2, 3	1, 2	1, 2
Insect/disease occurrence	0	3, b	2, 3, b	1, 2, 3, b	2, 3, b	2, 3, b	2, 3, b	1, 2, b	1, 2, b
Fire occurence	0	2, 3, b	1, 2, 3, b	1, 2, 3, b	1, 2, 3, b	1, 2, 3	1, ,2, 3	1, 2	1, 2
Forage production	3, a, b	2, 3, b, c	1, 2, 3, b, c	1, 2, 3, b	2, 3, b	2, 3, b	2, 3, b	1, 2, b	1, 2, b
Range condition	3, a, b	2, 3, b, c	1, 2, 3, b, c	1, 2, 3, b	2, 3, b	2, 3, b	2, 3, b	1, 2, b	1, 2, b
Range cover type	0	3, b, c	2, 3, b, c	2, 3, b	0	1, 2, 3	2, 3, b	1, 2, b	1, 2, b
Water courses	0	2, 3	2, 3, c	1, 2, 3	1, 2, 3	1, 2, 3	1, 2, 3	1, 2	1
Water bodies	3	2, 3	1, 2, 3, c	1, 2, 3	1, 2, 3	1, 2, 3	1, 2, 3	1, 2	1
Turbidity	0	2, 3, b	2, 3, b	2, 3, b	0	0	2, 3, b	1, 2	0
Water temperature	0	0	2, 3, b	0	0	0	0	0	0
Snow cover	3	2, 3	1, 2, 3	1, 2, 3	1, 2, 3	1, 2, 3	1, 2, 3, b	1, 2	1, 2
Geologic landforms	0	3, a, b, c	2, 3, a, b, c	2, 3, a, b, c	2, 3, a, b, c	2, 3, a, b, c	2, 3, b	1, 2, b	1, 2, b
Soil moisture	0	3, b	2, 3, b	1, 2, 3, b	2, 3, b	2, 3, b	1, 2, 3, b	1, 2, b	1, 2, b
Soil map units	0	3, a, b, c	3, a, b, c	2, 3, a, b, c	2, 3, a, b, c	2, a, b, c	2, 3, b	1, 2, b	1, 2, b
Cultural‡:									
Major highways	0	3	1, 2, 3	1, 2, 3	1, 2, 3	1, 2, 3	1, 2, 3	1, 2	1, 2
Secondary roads	0	0	2, 3	1, 2, 3	1, 2, 3	1, 2, 3	1, 2, 3	1, 2	1, 2
Logging roads	0	0	0	3	1, 2, 3	1, 2, 3	1, 2, 3	1, 2	1, 2
Jeep/hiking trails	0	0	0	0	0	2, 3	2, 3	1, 2	1, 2
Automobiles	0	0	0	0	0	1, 2, 3	1, 2, 3	1, 2	1, 2
Buildings/houses	0	0	0	3	2, 3	1, 2, 3	1, 2, 3	1, 2	1, 2

Recommended Level of Use

- 0 Not recommended for creation of data layer.
- 1 Recommended for small-area project where great detail is required (e.g., riparian mapping).
- 2 Recommended for medium-area projects where broader classifications are useful (e.g., district or forest).
- 3 Recommended for very large-area mapping projects where little detail is needed (e.g., State or country).
- a Used in conjunction with terrain data (slope, aspect, elevation).
- b Used in conjunction with field-collected data.
- c Used in conjunction with photo interpretation.

*Vegetation, ecology, geology, topography, soils, water, air/climate, wildlife/fish.
‡Constructed features, historic/prehistoric features, recreation setting, socioeconomic features.

***Note*:** The GIS data layers shown are a subset of the recommended USGS National Geographic Information Structure as well as other applicable data layers that may be derived from remotely sensed data.

Resource managers may need many additional data layers not listed in table 2b, including unique/critical habitat, riparian mapping, or fuels modeling. In some cases, these data themes may be derived from existing data and appropriate GIS or cartographic models. In other cases, it will be necessary to extract the data of interest directly from remote sensing imagery and ground measurements.

Other Georeferenced Data

Spatially referenced data sets include inventory and other measurement data referenced to a specific geographic location, as well as geographically referenced data created for cartographic and GIS applications that describes natural or manmade features of the environment.

Location information may be a part of the individual measurement record implicitly referenced through a separate file containing the coordinate locations or data collection stations in an explicitly referenced data set. Until the advent of GIS and Global Positioning System (GPS) receivers, there was neither the need nor the technology to cost-effectively determine the precise location of widely distributed field locations. Inventory plots, weather stations, and pollution monitors typically produce georeferenced data sets that are often available in a basic file structure.

Implicitly georeferenced data can be obtained by defining coordinate reference values for a record along with a geographic offset between successive measurement values in the record. Terrain data files are often structured in this way. Implicitly georeferenced information can also be obtained by providing a field that links the data to an established feature such as a county or census tract.

Digital data describing natural and manmade features of the Earth are used to support a variety of software applications. The most common applications in natural resource fields include a GIS and digital cartography. The strength of a GIS lies in its capability to analyze the relationships between map features. Digital cartography is used in data collection and editing as well as in producing fully symbolized maps.

Locations of Information

As a good manager of the Enchanted Forest, you should search for all probable sources of information. Be thorough. Information can be found in most land and resource administering agencies, remote sensing centers, libraries, agriculture statistical services, census bureaus, land-use institutes, industries, consulting firms, professional societies, planning departments, bureaus of statistics, universities, environmental organizations, research organizations, archives, intelligence and law enforcement agencies, the military, international groups, and public and private organizations that specialize in resource information. One source will frequently lead to another.

Still, obtaining copies of information may be difficult at times. Some information may not be published or may not be available for national security reasons. Data may not be readily available when the supply of a published map or inventory report is exhausted. The time and cost to reproduce electronic data may be significant, especially for an organization not funded to distribute data. At other times, people may not want to share their data because they fear being “scooped” or having the

information used in a negative manner. Cooperation toward some common goal is one way of getting around these kinds of barriers.

Where integrated information systems are available, data for recurring requirements will often exist within the system. In these cases, the user can proceed directly to the evaluation of the suitability and quality of the data. Integrated information systems, including GIS's, are likely to contain only a relatively small portion of the data available for an area of interest. Because of the expense of data entry and maintenance, only those existing data sets essential for frequently recurring applications are likely to be included in the system. Additional data sources may be needed to conduct a specific activity. In cases where an integrated information system is not in place or the requirement differs from the needs anticipated when the system was established, a search for data must be conducted to meet the requirements specified in the INA. To assure quick response to new requirements, an inventory of existing data for the area of interest should be developed.

When existing data cannot be located to meet a specific requirement, expand the search to include ancillary data from which the required information may be derived. For example, a slope map may be required, but may not be available. In this case, the user should expand the search to include elevation data from which to produce a slope map. If an existing elevation data set cannot be located, expand the search to include imagery from which the user can derive elevation and eventually slope information. Trade statistics, records of treatments or harvests, mapping updates, and repetitive remote sensing coverages are potential sources of change and trend information.

The process of searching for existing data must be separated from the process of evaluating the usefulness of the data. Selecting the combination of data sets that best meets the information requirement of a particular resource inventory or analysis activity is generally an iterative process. Potential information users must evaluate many factors to determine the data that best meet the requirements of the analysis, including information requirements, data availability, analysis procedures, accuracy requirements, costs, and timeliness. The scale of a State soils map, for example, may not be suitable for evaluating the erosion potential of a river basin. If it is the only data set available, however, it might serve as the basis for stratifying the area for ground sampling to acquire more site-specific information.

Do not limit your search to obvious locations. Take advantage of computerized data inventory and online information retrieval services. The USGS's National Mapping Program, for example, has set up Earth Science Information Centers nationwide to provide information on existing data. Information on a broad range of topics is available, including aerial photography and geologic, hydrologic, topographic, and land use maps. In some cases, the information can be ordered directly from a particular center. In other cases, the center may refer you to the organization holding the data. The center nearest your study site will usually be more familiar with data to meet your specific needs. This program also maintains a network of State cooperators who can provide additional help. The National Agricultural Library is another source of aid in locating existing information.

Many agencies maintain data centers with computer-based retrieval systems and staffed by personnel familiar with the peculiarities and limitations of specific data sets. Look for data from related sectors—not just from the obvious sectors. Agricultural research publications and census reports are two sources that natural resource managers commonly overlook.

The principal sources of digital data are organizations charged with conducting land management, regulation, or research activities. Data are available in agency-developed or nationally recognized interchange formats or in the interchange format of the software used to develop the data. The DLG is currently a standard format for digital planimetric, cartographic data. This format, although structured to capture much of the information content of a file, does not fully meet the requirements of current applications. The Forest Service is evaluating the successor to this format, Digital Line Graphic Enhanced, to meet the more stringent requirements of current systems and to provide greater flexibility within a single interchange format. The new Spatial Data Transfer Standard, as outlined by the USDI Geological Survey (1992a), became a requirement for transfer of data between Federal systems in early 1994.

All DLG data distributed by the USGS are DLG-Level 3, which means the data contain a full range of attribute codes, have full topological structuring, and have passed certain quality-control checks. The intermediate (1:100,000 scale) DLG data files that cover transportation and hydrography are available for all States except Alaska. Intermediate DLG data are sold in 30- by 30-minute units, which correspond to the east or west half of USGS 30- by 60-minute 1:100,000-scale topographic quadrangle maps. Each 30-minute unit is produced and distributed as four 15- by 15-minute cells, except in high-density areas, where the 15-minute cells may be subdivided into four 7.5-minute cells.

Data of varying resolutions may be available from various agencies (Federal Geographic Data Committee 1993). Although quality checks are applied to these data sets, users will often have to reformat and verify the data to meet their own needs. A specialist with subject area expertise and familiarity with the data collection systems may be required to perform these tasks. Geographical Information Systems vendors and advanced users are alternative sources for standard data sets. These organizations have already assembled and transformed the data into the required GIS format. If funds are available, data from these sources are well worth the cost.

Geographical Information Systems data sets are also available from organizations with responsibility for a specific resource or geographic region. For example, digital data for 16 percent of the National Wetlands Inventory are available from the USDI Fish and Wildlife Service (FWS), and soils data for selected counties and areas throughout the United States and territories are available through the USDA Natural Resources Conservation Service (NRCS) from its Soil Survey data base. State departments of forestry and natural resources or water management districts may have assembled GIS data themes suitable for specific projects (Warnecke and others 1992).

Sources of information:

United States:

- Federal Geographic Data Committee (1993)
- Warnecke and others (1992)

International:

- Forest Resources Division
  Food and Agriculture Organization of the United Nations
  Viale delle Terme di Caracalla
  00100 Rome, Italy

- United Nations Environment Programme
  Global Resource Information Database
  P.O. Box 30552
  Nairobi, Kenya

- World Forestry Institute
  4033 SW Canyon Road
  Portland, OR 97221

- World Resources Institute
  1709 New York Ave., N.W.
  Washington, DC 20006

Sources of models and procedures include reference and text books, symposia proceedings, journal articles, and research or administrative reports. The "Manual of Remote Sensing" (Colwell 1983) is the most comprehensive reference to the technology linking image interpretation with ground conditions. Patently helpful texts are listed in its bibliography. More than a dozen journals are available dealing with either GIS or remote sensing. Articles on the technology also appear in subject area journals. The National Agricultural Library provides literature searches for government support activities. The Canadian Centre for Remote Sensing's Online Retrieval System is the world's foremost bibliographic system and document collection in the field of remote sensing. Reviewing the appropriate literature can improve the efficiency of the process and avoid costly mistakes.

Documenting Information

Whatever the source of data, seek complete documentation to ease the process of evaluating existing information. Develop a description of the data at the time you locate them, and take the opportunity to question specialists familiar with the data and their sources. Full data descriptions will save valuable time during the assessment of data utility.

Documentation should allow users to trace the lineage of all documents and products back to a specific primary source. A primary source is an initial record of field or photogrammetric observations or measurements. A primary source may be a digital file (such as raw multispectral digital imagery, GPS receivers, or an electronic notebook), a map (if the map is produced in the field and measurements and

observations recorded directly on the map base), a log book or survey notes, or an annotated remote sensing image (such as an aerial photo or orthophoto). Secondary sources may be plot summaries or published reports and maps. Knowing the characteristics of the source materials and the changes and conversions that the information has gone through from its primary source to its present form provides insight into the positional accuracy, thematic content, and resolution of the data.

Types of data documentation to obtain:

- Source(s) of the original data and date and method of collection
- Scale(s) or intensity and resolution of the original data, including the area of smallest mapped unit or broadest sampling frequency
- Agency inventory programs that relate to the data and its limitations as perceived by the originator and users
- Significance or importance of the resource (or information) to the agency and the rationale for classification schemes or setting of priorities
- Quality control checks applied in data collection, compilation, and summary, and
- Name and telephone or fax number (or E-mail address) of a person to contact for further information.

(Based on Lund 1986b.)

Summary

Several kinds of information are commonly available, from personal knowledge through data from remote sensing. Primary sources of spatial data for Federal and State organizations are given in Federal Geographic Data Committee (1993) and Warnecke and others (1992). The key to a successful search for information is to be thorough and persistent. Ask questions—one source often leads to another. When gathering data, seek as much documentation as possible about the source, quality standards, definitions used, etc. This information will later be useful in determining the suitability and quality of your data for your particular purposes.

Chapter 4: Determining Data Utility

After locating available information, you need to determine if it is suitable for your purposes. If not, then you must collect new data.

As Enchanted Forest manager, you may find that existing information is difficult to evaluate because "information quality" means different things to different people. Confusion often arises because of divergent perspectives. When developing a corporate data base system in a complex organization, data utilization must be considered from several perspectives, including data base management, cartographic design, legal considerations, and scientific and analytic aspects (see table 3). Existing information may be good from one or more of these perspectives and poor from others. If an information system is to be a useful and trusted basis for supporting management decisions, then the data must stand up to legitimate challenges under each criterion from every perspective shown in table 3.

This chapter discusses the basic concepts, major classes of error sources, and specific issues that should be considered when evaluating data quality.

Table 3—Perspectives on resource information quality.

Perspective	*Source of criteria*
Data administration	Information engineering
	– Accessibility
	– Consistency
	– Documentation
	– Security
Cartographic	Cartographic convention/land surveying
	– Visual quality
	– Communicability
	– Mapping standards
Legalistic	Regulations, laws, directives
	– Specifications met
	– Consistent with law
	– Appeal to expert opinion
Scientific	Logical and mathematical rigor
	– Appropriate assumptions
	– Valid models
	– Veracious data
	– Significant results

Basic Concepts

There are basic concepts relevant to natural resource information that need to be understood if a discussion on evaluating data utilization is to be meaningful. These concepts include information and data, classification systems, suitability and quality, and standards of accuracy.

Information and Data

Usually we refer to data as some primary measurement and to information as processed data that summarizes the raw input data into a useful form; but more often than not, the terms are used interchangeably. In most applications for corporate data bases and GIS's, it is preferable to enter raw or basic data rather than information or interpreted information. One reason is that basic data can usually be reprocessed using newly developed models or applying new statistical summarization methodologies, whereas the final information cannot usually be so treated. Although this precept is sound in principle, it is not always possible in practice, given constraints of time, personnel, and funds for doing field inventories.

For broad-area analysis, it is practically impossible to acquire direct measurements of properties at high enough resolution to derive land qualities directly. For example, we could conceivably develop a forest type map if we enumerated every tree in the forest and listed its coordinates and species in a data base. However, we rarely would have the funds or the need to do this. Instead, we would stratify the land into broad vegetation classes using remote sensing and then use field samples to derive forest type classes. When considering whether to include an existing type map into a corporate data base, for example, both the reliability of the remote sensing interpretation and field survey data as well as the relevance of the information to present issues and operational objectives must be taken into account. Once such information is entered into the system, the user must know and respect the limits of its utility.

Measurements—Measurement data may come from a wide variety of sources, including field observations, surveys, and remote sensing. Included are data from field-deployed devices (such as rain gauges and fuel moisture sticks), data collected by field personnel (including tree diameters and heights), navigation and surveying system positions, and data gained by imaging and nonimaging remote sensing systems.

Derived information (data)—Derived information may come from measurements of one or more existing data sets by summarizing, classifying, and analyzing the data with models and processes. Tree volume (computed from tree diameter at breast height, or dbh) is an example of derived information calculated from measured data. Derived information may take many forms, including maps and statistical summaries.

The process of deriving information involves a decisionmaking process that is linked to its application. The process by which the information was derived should be understood and documented. Aerial photography is an example of a remote sensing measurement data set. Photogrammetrists extract information regarding location and description of specific features from the imagery to provide information

for products such as type maps, land form maps, and USGS 7.5-minute, 1:24,000-scale quadrangle maps. The development of the map product involves a well-defined process that includes selection of features, generalization of positional location, and classification or symbolization of the features.

Information-deriving processes employ diverse techniques, including *manual and computerized systems* to extract information from remote sensing imagery, *statistical procedures* to summarize forest inventory measurements, and *cartographic models* to estimate the effects of management actions or system behavior.

Transformation is a subcategory of each of the above processes. Its objective is to enhance the usefulness of the data without modifying the information content; that is, it is content neutral. One can transform either measurement data or derived data products. Transformation is often the initial step in creating data products. Examples of transformation include conversion from English to metric units and rectification or geocoding of remote sensing imagery. A derived map product in a GIS is transformed when the data are projected between coordinate systems or during transfer to a GIS file structure. While we want transformations to be neutral, actions as simple as rounding or truncating significant figures could bias the outcome.

Disclosing the nature of transformations applied to a data set helps users evaluate data quality. Although the measurements may be modified by transformation, they still represent values of the original variables. Transformation of satellite imagery to a geocoded base suitable for use in a GIS requires subtle modification of the image reflectance values that may affect performance of automated classifiers. However, the digital values in the image still represent the intensity of reflected energy in specific wavelength bands.

One aspect of professional practice in natural resources is the knowledgeable interpretation of data and the conversion of various data into information useful for people. Professionals routinely translate time and place data into useful information with known estimates of precision and reliability. Consumers of resource information often view interpreted products generated through a GIS with a sense of precision and reliability greater than the data warrant. We have no convenient way to display the relative precision and reliability of the various layers or types of natural resource data that are compiled in a common data pool or geographic display. Mapmakers, through time, have used scales and legends to communicate the relative precision and reliability of the information displayed to the map user. Our electronic maps or GIS's often mix data of vastly differing levels of detail, as well as data of varying precision and reliability. The interpreter must decide if the information is sufficient to fulfill the needs. As discussed above, the conclusion that the information is sufficient to meet the needs has a scientific and "professional" component as well as a social or political element based upon the issue of the day.

Models usually take the form of equations to transform some type of measured or observed data to some kind of predicted information. For example, a model to predict tree volume may use measured variables of dbh and total tree height times a coefficient for a particular tree species. There could be several equations used for a

given species in a given area. The user should review the models and transformation processes to determine which are most applicable for the use at hand.

The INA specifies the information products necessary to support resource management, and gives a general description of the data, models, and procedures that must be used to develop the data product. When searching for existing data, one should search for examples of the models and procedures, in addition to the data themselves. Traditionally, we rely on established procedures defined in such documents as agency handbooks to gather and summarize data. In our current situation, the development of a wide variety of information products with only general guidance taxes the inventory specialist.

The Forest Service National Riparian Initiative is one of many activities and emerging issues that generate requirements for specialized information products. The initiative describes a requirement for inventorying and monitoring riparian areas; however, the specific approval and definition of products are left to local decision. Mereszczak and others (1990) present four case studies that employ remote sensing technology ranging from high-altitude aerial photography to satellite imaging in meeting the requirements of the initiative.

Existing models may have to be adapted to specific situations. As existing measurements and data layers improve the timeliness of information products, the use of existing models and procedures can improve the likelihood of success, reduce costs, and provide a more definable, scientifically valid product.

Classification Systems

Classification is the process of grouping features or measurements. The technique may be applied to both continuous measurements and categorical data. Classification may take place during data collection, storage, or analysis.

Two motives for classifying information are to reduce the amount of data (preclassification) and to enhance communication (postclassification). In preclassification, objects are classified either to reduce the amount of data needing storage and analysis (by grouping objects into categories based on statistical summaries and discarding specific measurements) or to reduce the amount of data needing collection (by stratifying data to increase the efficiency of a survey). In postclassification, we can enhance communication by first understanding that there is a limit to the amount of information that can be comprehended at one time. So we often group objects into categories to draw attention to patterns or trends that we want to emphasize. For example, forest land is often separated from other lands when we wish to address forest-related subjects. We may also divide forest land into areas with forest cover and those that have recently had forest cover removed in order to emphasize the impact of deforestation.

The reduction of measurements and naming of features resulting from classification can enhance our understanding of a complex process or environment, help communication, and enhance the decisionmaking process (Burrough 1986, 1989). A satellite image, for example, consists of rasters of measured reflectance values sampled at specified bands in the electromagnetic spectrum. We can visually perceive the location of features in this complex measurement set, but classification

of the image into discrete land cover classes is essential if we are to use the data to identify inventory strata. Continuous slope values may be grouped into classes as input into a harvest suitability model. In each of these cases, the original measurement values are retained, ensuring that the data can be reclassified if another grouping of the measurements is more appropriate for a future application.

Data are sometimes classified in the process of automating the information and entering it into a GIS or DBMS. Categorical data may take less than one-fourth the storage space of continuous measurements. But reduced data storage cost is rarely sufficient justification for classifying continuous measurement data (Burrough 1986). Data from individual inventories may also be classified to fit a standard data structure established at a higher organizational level.

Classification is a significant issue in the design and implementation of inventorying, mapping, and monitoring activities. The impacts of classification systems used in data collection are especially far-reaching, because there is no way to reverse the process and extract measurements from data classified during data collection. The level of classification-specified data collection in an inventory or mapping project can significantly affect the cost of the project. In designing an inventory, the reduced costs and improved efficiency of classifying data during collection must be carefully weighed against their impact on the information requirements of the inventory.

It is difficult to establish the appropriate level of classification in data collection for multiresource survey, and it is impossible to gauge its effect on the utilization of the data to meet future requirements. In conducting timber inventories, for example, it is common practice to record tree diameter measurements in 2-inch (5-cm) increments and height measurements to the nearest full log (16 feet or 5 m). While this classification of measurement during data collection may have little impact on inventory volume estimates, it will significantly reduce the utility of the data for constructing timber volume tables.

Classification also helps the identification of discrete geographic units in stand mapping and similar activities. Timber stands, for example, are often delineated based on ocular estimates of the proportion of species groups within the stand. This practice can result in arbitrary delineations that are of relatively little value for GIS modeling applications. This is especially true for stands that fall close to the class limits. Stand classification procedure used by the Forest Service in the Southern United States defines a hardwood stand as one where more than 70 percent of the dominant and codominant crowns are hardwood and a hardwood pine stand as one where 51 to 69 percent of the dominant and codominant crowns are hardwood. The class limits are continuous, but only 1 percent in hardwood crowns separates the two classes. This type of stand map is of relatively little value, for example, in evaluating the accuracy of satellite-derived land cover classifications.

The impact of the process on data utility should be carefully considered before the process is applied in data collection and automation. The level and structure of classification in existing data are important factors in evaluating data utility. When both original data and classified data are available, it is generally better to obtain the

unclassified data. In this process, however, and in designing an inventory, it is important to weigh the loss of utility against the cost of reclassifying and verifying the data to meet current requirements.

Suitability and Quality

The utility of a candidate data set to meet the specific requirements of an information need is dependent on the suitability and quality of the data set. Suitability describes the applicability of the data set to a specific requirement. Quality defines how closely the data set conforms to some type of standards. Quality may be expressed by accuracy, the likelihood that a value or prediction will be correct (Aronoff 1989). Because the process of assessing data quality requires a more detailed examination of the data set than determining its suitability, we must separate the two parts of data utility and assess suitability first. Quality assessments then must be performed on those data sets that meet suitability criteria.

Let us suppose, for example, that our goal is to determine the potential soil loss in a watershed from sheet erosion on the Sharewood Forest. In developing the model, we determine that the location of recent clearcuts is an important factor. In conducting the data inventory, we locate three data sets that have the potential of meeting this requirement. The first data set is a stand map created 10 years ago. The second is a stand map created last year, and the third is a Thematic Mapper (TM) satellite imagery acquired last year. Based on the literature from which the model was developed, three data suitability criteria are defined:

- First, the analysis must include all cutting that occurred within the past 3 years.

- Second, it must include clearcuts 50 acres (20 ha) or more in size.

- Third, because the land base of the Sharewood Forest is intermixed with private land, our mapping must be very accurate. At least 90 percent of mapped features must be within 10 percent of their true area, centroids of the features must be within 164 feet (50 m) of their true locations 90 percent of the time.

Because recent clearcuts are defined as those less than 3 years old, we reject the first data set as not suitable for the current analysis. The 10-year-old stand map might, however, be suitable for an analysis of the trend in the size of clearcut units. Both the 1-year-old stand map and the TM data set pass the first data utility criterion. Through inspection of the 1-year-old stand map, we find that it shows clearcuts of 10 acres (4 ha) in size or greater. A review of the literature determines that it is possible to meet the second utility requirement with classifications of TM imagery. Therefore, this data set also meet the data suitability requirements.

Next, the quality of the two data sets that have met the suitability criteria is tested. In this process, random points on the type map are compared to base photography flown the same year. More than 30 percent of the clearcuts of less than 50 acres (20 ha) are not delineated on the stand map. Although we accepted the stand map based upon data suitability, we reject it based on inadequate data quality. Because the TM data form a measurement data set, an evaluation of data quality is deferred

until after the imagery is processed and a map of clearcuts derived. We can, however, incorporate the data quality requirements into the classification and accuracy assessments that would normally be conducted in deriving GIS coverages from satellite imagery.

Accuracy Standards

The data used by natural resource managers provide a description and, in georeferenced data, the location of a feature or class of features in the natural environment. The creation of accurate data requires the producer to follow a well-defined process. Accurate data will result when trained individuals collect the information using appropriate tools, technology, techniques, and procedures that include verification and quality control.

A statistical measure of accuracy is built into the design of many resource inventories. Error terms that define the accuracy of the data are computed as part of the estimation process. The level of accuracy of satellite image classifications of land cover can be inferred from confusion matrixes computed from the ground verification data.

Spatial data used by natural resource managers can be divided into two classes, base data and resource data. Base data are created primarily by mapping agencies, and resource data primarily by resource management agencies. Producers of base data establish objective accuracy standards, whereas resource agencies rely more on process and subjective measures to define product accuracy. Because data from resource agencies are more widely shared among government organizations and available to the public, the trend is toward the establishment and adherence to objective standards for resource data. Resource management agencies are required to provide the public with copies of their data, including GIS data layers. In the near future, agencies will have to provide both data and metadata (data about data) describing their resource information. According to Ogrosky (1992), "The Federal Geographic Data Committee draft metadata standards specify eight data quality elements treating positional accuracy, attribute accuracy, data model integrity, and completeness (capture criteria)."

Base data include planimetric, cultural, hydrographic, hypsographic, and elevation data. In the United States, these data are provided primarily by the USGS and prepared either directly by USGS or in cooperation with another government organization or private contractor. The data may subsequently be modified by agency mapping centers to meet their specific requirements. Mapping centers such as the Forest Service Geometronics Service Center (GSC) and the NRCS National Cartographic and GIS Center meet agency-specific requirements. These organizations have adopted procedures and standards similar to those of the USGS. Procedures used by the USGS are designed to ensure that printed maps meet National Map Accuracy Standards (NMAS) (American Society of Civil Engineers 1978). A draft revision (USDI Geological Survey 1992b) of the NMAS, currently under review, includes separate estimates of horizontal and vertical accuracy (X,Y,Z), and describes testing procedure and the labeling of products. An effort is being made to ensure that the revised NMAS can be applied to both printed and digital data.

In addition to planimetric data in printed maps and DLG's, the USGS produces raster elevation data sets. Vertical accuracy of digital elevation model (DEM) data is dependent upon the spatial resolution (horizontal grid spacing) quality of the source data, collection and processing procedures, and digitizing system (USDI Geological Survey 1990). The USGS has identified three levels of accuracy for standard 7.5-minute DEM's. The desired vertical accuracy level for Level 1 products (this includes all currently available DEM's) is 23 feet (7 m) with a maximum acceptable level of 49 feet (15 m) and an absolute elevation tolerance of 164 feet (50 m). The standard 7.5-minute DEM header includes fields for accuracy information.

The accuracy of data sets developed on a national basis by resource agencies is dependent on the source material and procedures used in their development. County soil association maps produced by the NRCS and national wetland maps produced by the FWS are examples of this type of product. The accuracy of these products is generally defined in terms of goals and processes rather than absolute standards. To understand the accuracy of these products, the users must understand the processes used in their production.

Stringent accuracy standards have not generally been established for data intended primarily for use within the agency. While no absolute standard has been established, for example, for stand mapping by the Forest Service in the Southern United States, a level of product accuracy is developed and maintained through training, adherence to procedures, and review of completed prescriptions. As these data are integrated into GIS's and improved technology becomes available, the accuracy of these products will increase.

Time, space, and relative accuracy—Political boundaries and land survey data are one way of demarcating land to show who owns what. They rarely, however, correspond to boundaries that have biophysical meaning. We can assign most of the natural resource data we use in land management to a place on a map. Natural resource areas often have an implied time value associated with them. For example, foresters assign a rate of growth to a particular stand of trees as an indicator that the data used today will have a different value later. On the other hand, hydrologists describe river flow as a variable through time, with estimated return frequencies of high and low flows. Various disciplines concerned with natural resources view the Earth through quite different temporal lenses. Different resources may have quite different and acceptable accuracies associated with them; indeed, even a single resource has different relative accuracies associated with different end uses. Timber surveys used to develop forest plans may require less accurate information than soil surveys of the same area that are used to implement the results of forest planning. Similarly, timber data gathered for a timber sale usually need to be more accurate than those used for State or national assessments.

The human element—One professional role in natural resources is the knowledgeable interpretation of data and the conversion of various data into information useful to people. Professionals routinely translate time and place data into useful information with known estimates of precision and reliability. Consumers of resource information often view interpreted products generated through a GIS with a sense of precision and reliability greater than the data warrant. We have no convenient way

to display the relative precision and reliability of the various layers or types of natural resource data compiled in a common data pool or geographic display. Mapmakers, through time, have used scales and legends to communicate the relative precision and reliability of the information displayed to the map user. Our electronic maps or GIS's often mix data of vastly differing levels of detail, as well as data of varying precision and reliability. The interpreter must decide whether the information is sufficient to fulfill the needs. As discussed above, the conclusion that the information is sufficient to meet the needs has a scientific and "professional" component, as well as a social or political element based upon the issue of the day.

Major Categories of Error Sources

Table 4 shows common sources of error in using a GIS.

Table 4—Common sources of error in corporate data bases.

Stage	*Source of error*
Data collection	Field data collection
	Existing maps or overlays used as source data
	Analysis of remotely sensed data
Data input	Data entry
	Arbitrary geographic feature (e.g., edges of vegetation types that do not actually occur as sharp boundaries)
Data storage	Numerical precision
	Spatial precision
Data manipulation	Class intervals
	Boundaries
	Models, processes, and overlay procedures
Data output	Reporting or scaling of overlay procedures
	Output device
	Medium
Use of results	Comprehension of information
	Use of information

Source: Modified from Aronoff 1989.

Evaluating existing information requires an understanding of the sources and consequences of errors. Below, we have grouped the sources of errors in table 4 into three categories: source (or inherent) errors; processing (or operational) errors; and use (or modeling) errors. This grouping provides a useful framework for evaluating existing information.

Source Errors

Source errors inherent in the data before processing include:

Positional error—The accuracy of base series maps, such as 1:24,000, constrains forest and strategic project level work.

Mixed and unknown spatial resolution—Resource mapping of large areas often reveals large variations in the size of mapping units.

Heterogeneity—Land cover classes are entered in a geographic data base as closed polygons, and attributes are assigned to them (i.e., a class name). Variation within the polygons is unknown. Variation cannot be disaggregated then. Problems arise when several layers of preclassified data must be integrated to characterize particular land units.

Inappropriate classification systems—Sometimes older classification systems simply are not relevant.

Inadequately specified classification systems—When classes are poorly defined, interpreters may get significantly different results when using them.

Misinterpretations—Even when the classification system is relevant and well defined, if analyses are not controlled, different individuals may interpret classification criteria differently, and different remote sensing algorithms can result in vastly different results.

Incomplete data sets—Failure to delete outdated information or incorporate new information is a common problem.

Blunders—Accidental errors resulting from carelessness or mistakes are difficult to detect. Most blunders occur during the data preparation, collection, and entry phases of the project. Fewer blunders are likely to occur when personnel are well trained, follow appropriate procedures, and use high-quality source material. Automated logic and consistency checks assist in locating blunders.

Map preparation—Errors may result from generalizations and changes made to the data during manuscript preparation and drafting. Such changes include intentional displacements in the locations of lines to avoid slivers and to improve the visual quality of map displays.

Information on source errors is usually lost long before data enter the system. Therefore, the errors are difficult to model adequately. Source errors are generally more significant than errors introduced by processing. Therefore, methods should be devised for testing data for source errors before deciding to prepare the data for entry into the system (Goodchild and Wang 1988).

Without documentation on source error, the evaluation should be structured so that the most limiting errors are caught first. Even with documented data, we should confirm that the data meet the specifications they purport to.

Processing Errors

Errors may be introduced during machine processing and data transformations. Such processing errors include generalizations, modifications, and blunders introduced during coding attributes; digitizing; data conversions; and data base operations such as overlay and interpolations. Some of these errors may be effectively modeled and tracked through processing. However, some commercial GIS's do not include comprehensive error analysis procedures. Evaluate all digital data sources for the kinds of errors that may have been introduced at every stage of processing (a complete lineage of digital data should be known). Do not use digital data unless their source is known and can be evaluated for errors.

Use Errors

Use errors arise from using data for an application for which they are not suited. Misuse of data often results from mixing data that are incompatible with scale or resolution. Modeling error, a type of use error, has two components—specification and measurement error. Specification error amounts to using the wrong set of variables. Measurement error can be attributed to erroneous measurements on variables and to the erroneous calibration of coefficients.

Specific Issues

There are data suitability and quality issues specific to nonspatial and spatial data bases. Some of these issues are quite obviously raised by the georeferenced locations in spatial data and the general broad summary nature of many nonspatial data bases.

Nonspatial Data Bases

Nonspatial data bases include models and inventory reports.

Models—All models are simplifications of reality and involve some kind of generalization. Generalizations can be made through human interpretation (primarily subjective) or by quantitative analysis (primarily objective). There are advantages and disadvantages to both approaches.

Subjective generalization takes advantage of the tremendous capacity of the human mind to quickly synthesize information. Unfortunately, subjective generalizations are difficult to replicate because of differences between individual interpreters, which makes it hard to integrate information over space and time. It is also time-consuming and costly to transfer information processing abilities from person to person. It is even difficult for individuals themselves to remain consistent over time, because perceptions change with experience, and attitude and interest fluctuate.

Objective procedures (i.e., procedures that one programs) do not entirely eliminate subjectivity because someone designs the procedures. However, objective generalizations tend to be much more repeatable, and statistical tests can be devised to assure a level of repeatability. Criteria are expressed more explicitly and are more rigidly adhered to. However, it is very difficult to develop effective procedures that match the capacity of the human interpreter to synthesize information. Quantitative generalizations tend to require far more data than do subjective generalizations. Although criteria may be more explicit, they are often arbitrary and rely on acceptance of unwarranted assumptions. The primary danger of quantitative generalizations is that if data requirements are not met and if coefficients are not based on

sound analysis, results can be erroneous even when the formulation is correct (essentially, valid but untrue). Geographic modeling to support resource management decisions will always involve a mixture of subjective and objective procedures.

To determine if a data set is suitable for a particular application, describe the set of spatial models you are most likely to use. State the assumptions of those models and what the consequences of violating them are. This will help develop a list of potential problem areas that you will need to check.

Assess data suitability and quality needs in relation to the kinds of models the data are to support. Models include those designed for spatial interpolation (contouring and tessellations), input/output, growth, dispersion, and gravity.

It is important to differentiate between deterministic and stochastic models. Deterministic models show how certain variables interact in abstract space. They ignore errors in the data and uncertainty in specifying and estimating model parameters. Deterministic models often rely on unwarranted assumptions about reality, such as normal distributions, homogeneity, undifferentiated plane, and frictionless space. Although deterministic models are useful for instructional purposes, they can lead to serious misrepresentations when applied to real-world situations.

Stochastic models, on the other hand, try to account for data error, natural variation, and uncertainty. These models employ techniques that incorporate knowledge of uncertainty and error into the modeling procedures. Running such a model often provides a range of results that one may expect, given the uncertainty associated with the class assignments. Procedures such as this provide a basis for assessing the sensitivity of the models to the uncertainty known to be in the data. Similar procedures may be used to assess the impact of spatial error on the outcomes of models. Point or line locations can be chosen from a frequency distribution representing the known positional error in the data.

Inventory reports—Existing data, such as those found in FIA reports, may be suitable for establishing a corporate or GIS data base. Potential problems with existing data include level of aggregation, timeliness, completeness, form of collection, and the more subtle confusion of names with concepts used in the wider scientific community.

Often, information in published documents is too general for what is needed at the local level. Access to and use of the original plot level data may be more appropriate. Even these data, however, may not provide information for all cells in the GIS data base. In fact, they may provide only a single entry estimate, which would not allow one to calculate the error of estimate.

While you may find data readily available, you should assume that some updating is necessary. Suitable information can often be constructed from existing inventories and auxiliary information, such as remotely sensed classification of land class. Some of the needed data may not be directly available from inventory reports in any form. Finally, some data can be eliminated because of their poor quality.

Concerns for the quality of data must begin with a consideration of their statistical properties. If the report contains no error statement, one may assume that the error is as large as or even considerably larger than the estimate (the reported value) itself. In other words, the estimate may really not be significantly different from zero!

If there is some approximation of variance of estimates for some cells or polygons in the data base, it may be possible to establish preliminary information. Estimates of variances from cells with known variance may well serve to extrapolate to adjacent or similar cells in the data base. Inventory reports often include a description for estimating variances for subsets of the data. At some point, the lack of information about the statistical quality of the data will dictate the collection of new data for a given variable, regardless of its apparent existence.

The time value of data should not be overlooked, because the age of information has considerable importance when setting up a GIS data base. It is easy to grasp that data that is 30 years old is probably inappropriate for inclusion in a data base that is to display current information. However, the older information may have value for trend or historical studies. Similarly, for many data types, the lapse of a single year since measurement is seldom reason for rejecting the data. Obviously, there are methods to update inventory data. However, you will have to collect new information eventually.

Spatial Data Bases

Spatial data bases include maps, thematic overlays, and remote sensing. Issues in this group include scale, resolution, and coverage. Scale and resolution are complex concepts with multiple meanings, and both influence the results of any geographic analysis. Scale and resolution are thus critical variables to consider when evaluating the suitability and quality of existing data. Know the various meanings of scale and resolution and how they relate to each other. When establishing coverages, make sure you are aware of the three types of coverages (point, line, and polygon). It is also important to be aware of the type of polygon coverages (homogeneous versus heterogeneous).

When evaluating data for conversion, determine the utility of data and be sure that the original manuscript includes the reference coordinate locations and map projection data necessary to transform the data to ground coordinate values. Some data sets, such as many county soil surveys, may appear to be maps. In reality, they may be uncontrolled plat mosaics which, in turn, are extremely difficult to register to a map projection.

Scale—In the context of geographic information, scale has six meanings: mapping scale; positional accuracy; level in a categorical hierarchy (specific to general); level in a systematic or organizational spectrum (simple to complex); measurement scale (computationally adequate to inadequate); and level in a spatial hierarchy (small area to large area). There is often an implicit assumption that these various dimensions of scale (particularly categorical, systematic, and spatial) vary together in well-defined and predictable ways. However, this is not always the case. When evaluating existing resource information, consider each dimension of scale independently.

Mapping scale—Scale can be expressed as a representative fraction shown as 1/2,000 or 1:24,000. On a map with a scale of 1:24,000, one unit on the map represents 24,000 units on the ground. Map scales may also be described in the form of an equivalence of map to ground units. The equivalence for a 1:24,000 map would be 1 inch to 2,000 feet. Relative positional accuracy is often implied from map scale.

Maps compiled at a larger scale generally have higher positional accuracy than smaller scale maps. For example, the NMAS for maps with a publication scale of 1:20,000 or smaller require 90 percent of well-defined points to be within 0.02 inches (0.05 cm) of their true location (American Society of Civil Engineers 1978, Department of Defense 1981). For 1:24,000- and 1:100,000-scale maps, this translates to positional accuracies of 40 and 166 feet (12.1 and 50.6 m), respectively. It is poor practice to infer spatial accuracy from data map overlays or plots that have been digitally or photographically enlarged. USGS 15-minute quadrangle maps have been photographically enlarged from 1:62,550 to 1:24,000, for example, to provide an interim base map. These enlarged maps have the 104-foot (32-m) positional accuracy of the smaller scale map, not the 40-foot (12-m) positional accuracy of the 1:24,000 quadrangle.

Resource data themes (such as timber stands and soils) do not usually maintain the positional accuracy of the base maps on which they are delineated. It is difficult for field personnel to precisely transfer features located on the ground or on aerial photographs to the base map. The manual transfer or simple optical instrument (zoom transfer scope) techniques used by field personnel do not permit delineations to be precisely transferred to maps, especially in steep terrain.

Both the scale of the overlay or map and the scale of the source material from which it was derived (other maps, aerial photos, digital remote sensing data) are important in determining the positional accuracy of resource data. For example, stand boundaries depicted on a 1:24,000 base may have been transferred from either 1:12,000 or 1:60,000 aerial photography or could have been hand-transferred from a smaller scale map base. When evaluating existing mapped data, know the original source of the data and the method of transfer. You cannot infer the positional accuracy and mapping resolution of resource data from the scale of the base map on which they are depicted.

Positional accuracy—The scale at which a cartographic product meets NMAS is another representation of scale. Presently, there is no generally accepted positional accuracy standard for natural resource mapping. Acceptable levels of error may vary by application. However, since many Forest Service resource inventories and analyses have 1:24,000 mapping as a base, positional accuracy standards will limit maps of resource distributions.

Levels in a categorical hierarchy—Scale is often used in reference to a level in a categorical hierarchy. For example, biologists may study organisms at the species level or at more general levels of genera, orders, or phyla. Similarly, in soil surveys, one may map soils at the scale of soil series or at the scale of soil order. In the case of categorical hierarchies, the term "scale" refers to movement along a continuum from the specific to the general.

Levels in a systematic or organizational spectrum—Scale is also used in reference to a level along a systematic spectrum. Ecologists, for example, may conduct studies of biosystems along a spectrum of organizational levels, such as populations, communities, and ecosystems (Odum 1971). Generally, this spectrum represents a progression from lesser to greater complexity.

Measurement scales—There is a hierarchy of measurement scales based on their power in quantitative analysis: nominal, ordinal, interval, and ratio (Johnson 1993).

Nominal classes. Nominal classes represent categories with no particular order. Usually, these are characteristics that are not associated with quantities or quantitative measurements, such as soil type, vegetation type, or political area. Distinctions between classes are qualitative. An example would be land-use class (urban versus rural).

Ordinal classes. Ordinal classes are those that have a sequence, such as "poor, good, better, best." An ordinal class numbering system is often created from a nominal system in which classes have been ranked by some criteria. Ordinal measurements can be characterized by "greater than" (>) and "less than" (<) relationships between classes. Examples are:

- City classification: small–medium–large
- Terrain classification: plain–hill–mountain
- Stand-size class: nonstocked–seedlings–saplings–poles–sawtimber
- Population density: low–medium–high

Interval classes. Interval classes have a natural sequence like ordinal classes, but the distance between each value also has meaning. Numbers are used to describe classes, but the numbers do not have absolute value—the zero point in the scale used is arbitrary. For interval scaling, some type of standard unit is used, and the amount of difference between values is expressed in terms of that unit. Examples include elevation differences in units of feet or meters, and temperature differences in units of degrees centigrade or Fahrenheit.

Ratio classes. Ratio classes differ from interval classes only in having a natural zero point. Most measurements pertaining to length, area, and volume are ratio measures. Examples are:

- Elevation above a datum point in meters or feet
- Depth of snow or rain in inches
- Volume of streamflow in cubic feet per second

When evaluating existing information, consider whether the measurement scale of the information is appropriate for its intended uses. There are rules dictating what kinds of operations are allowable for nominal, ordinal, interval, and ratio data. Measurement scale is a key concern in assessing existing information. Many user errors arise from using mathematical operations on data types for which they are inappropriate.

Levels in a spatial hierarchy—The word "scale" is often used to describe the spatial extent of a study. Generally, a small-scale study is one confined to a small area (site-specific), whereas large-scale studies are extensive (covering broad geographic regions). As the spatial scale of a study increases, mapping scale (for practical reasons) decreases. An example of a spatial hierarchy is a large watershed, such as the Old Man River Watershed in the Emerald Kingdom (figure 4). The first-level watershed drains into salt water or a basin with no outlet, in this case the Misty Sea. Second-level watersheds, such as the Enchanted and Deep Dark Rivers, drain into the first-level watershed, the Old Man River. Third-level watersheds, such as Story Brook, drain into second-level watersheds. Each watershed is generally fanshaped, with the point of the fan being the mouth of the watershed and the rays of the fan being its tributaries. The fans become smaller and smaller as they progress through the hierarchy. The hierarchy contains numerous levels, and the level of concern depends on the scale of a project. A local project may occur only in a third- or fourth-level watershed, whereas a regional project might encompass a first-level watershed, including all levels in the hierarchy.

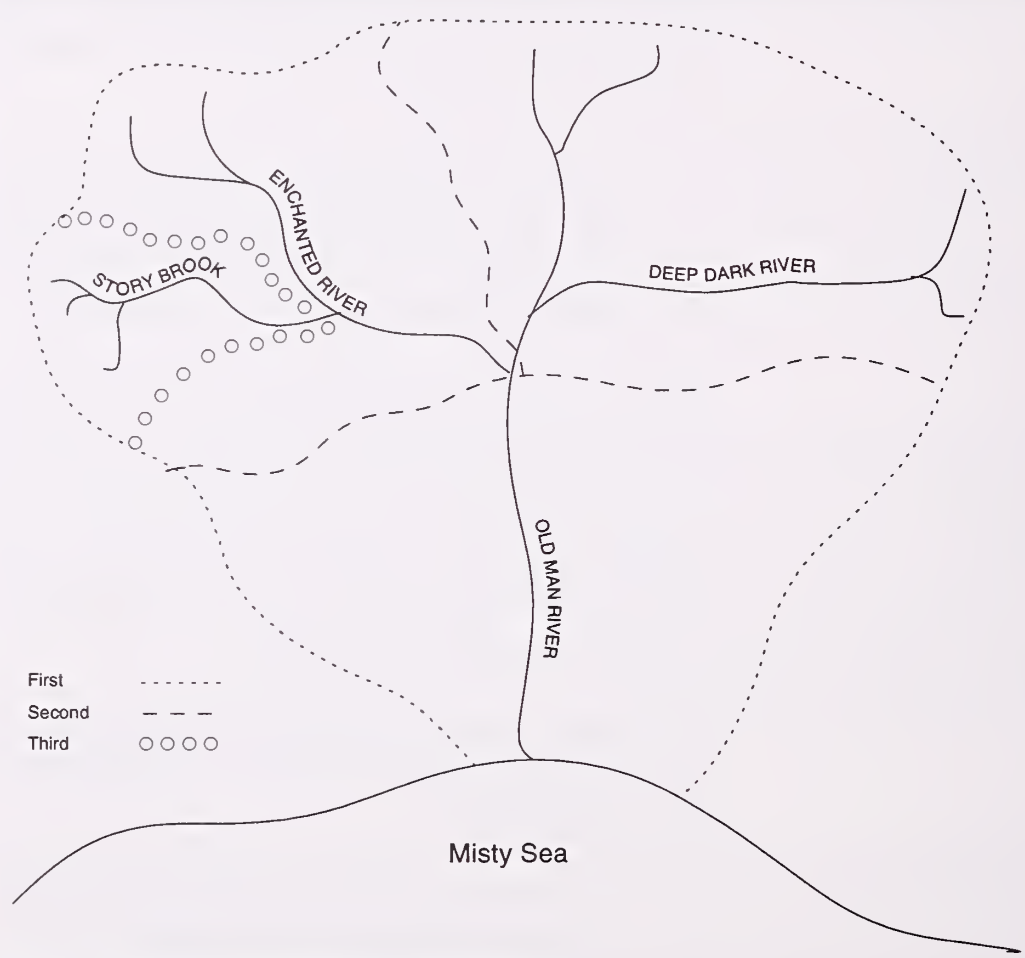


Figure 4—Levels in a spatial hierarchy of watersheds.

Resolution—Resolution refers to the level of detail required by an analysis or inherent in a data source. Four kinds of resolution are pertinent to natural resource information:

- *Categorical resolution*, or the number of categories in a classification system

- *Sensitivity resolution,* or the measure of how fine measurements or interpretations need to be to distinguish between classes

- *Temporal resolution,* or the time frame over which successive measurements are taken, and

- *Spatial resolution,* or the smallest discernible unit, often expressed in aerial photography as pixel size or number of line pairs per unit of distance area. Scale and spatial resolution are related in complex ways, and the relationships are not straightforward. Figure 5 shows aerial photographs taken of the same area with the same camera system, but at three different altitudes. While the resolution of the camera system remains constant, the features that an interpreter can discern change.

Spatial resolution varies, depending on whether one is referring to a map or imagery. Since thematic maps are generalizations of a portion of the Earth's surface and images show what is actually there, we can generally assume that there would be more detail in a photograph than in a map of the same area at the same scale. For example, a map may show forest stands, whereas individual trees may be discernible on aerial photographs.

Spatial resolution may also vary depending on the method used to develop a map. Unlike automated classification systems, mapping involving human interpretation depends more on subjective factors than on image resolution. We might expect a thematic map created through human interpretation to have larger polygons than one created by automated techniques, unless the same level of minimum mapping units is specified. Spatial resolution can also vary greatly as a result of the algorithms used in digital image processing.

Usually, there are tradeoffs between kinds of resolution. For example, to increase temporal resolution (get imagery more often), it is usually necessary, for economic reasons, to accept lower spatial resolution (fly at a higher altitude). The choice of film emulsion also affects sensitivity resolution (see figure 6). Color infrared film, for example, provides greater spectral sensitivity (ability to distinguish between moist and dry vegetation) at the expense of the superior spatial resolution of panchromatic black and white film.

Coverages**—**The basic types of cartographic coverages are point, line, and polygon. All may be associated with nominal, ordinal, interval, or ratio attributes, or any combination of them. A point location, for example, may carry the nominal attribute "campground," an ordinal attribute indicating degree of use (high, medium, or low), and numeric (interval or ratio) attributes such as acreage or number of campsites.

Similarly, one may associate lines with any combination of attribute types. For example, a line may carry the nominal attribute "trail," an ordinal attribute indicating degree of difficulty (hard, medium, or easy), and a numeric value for the trail's distance.

Polygon coverages are more complex, and their delineations fall into three general categories: administrative (political subdivisions, such as ranger district, management compartment, and census tract), arbitrary grids (rectangular or triangular), and natural variation (vegetation types, soil types, geomorphic structures, and elevation

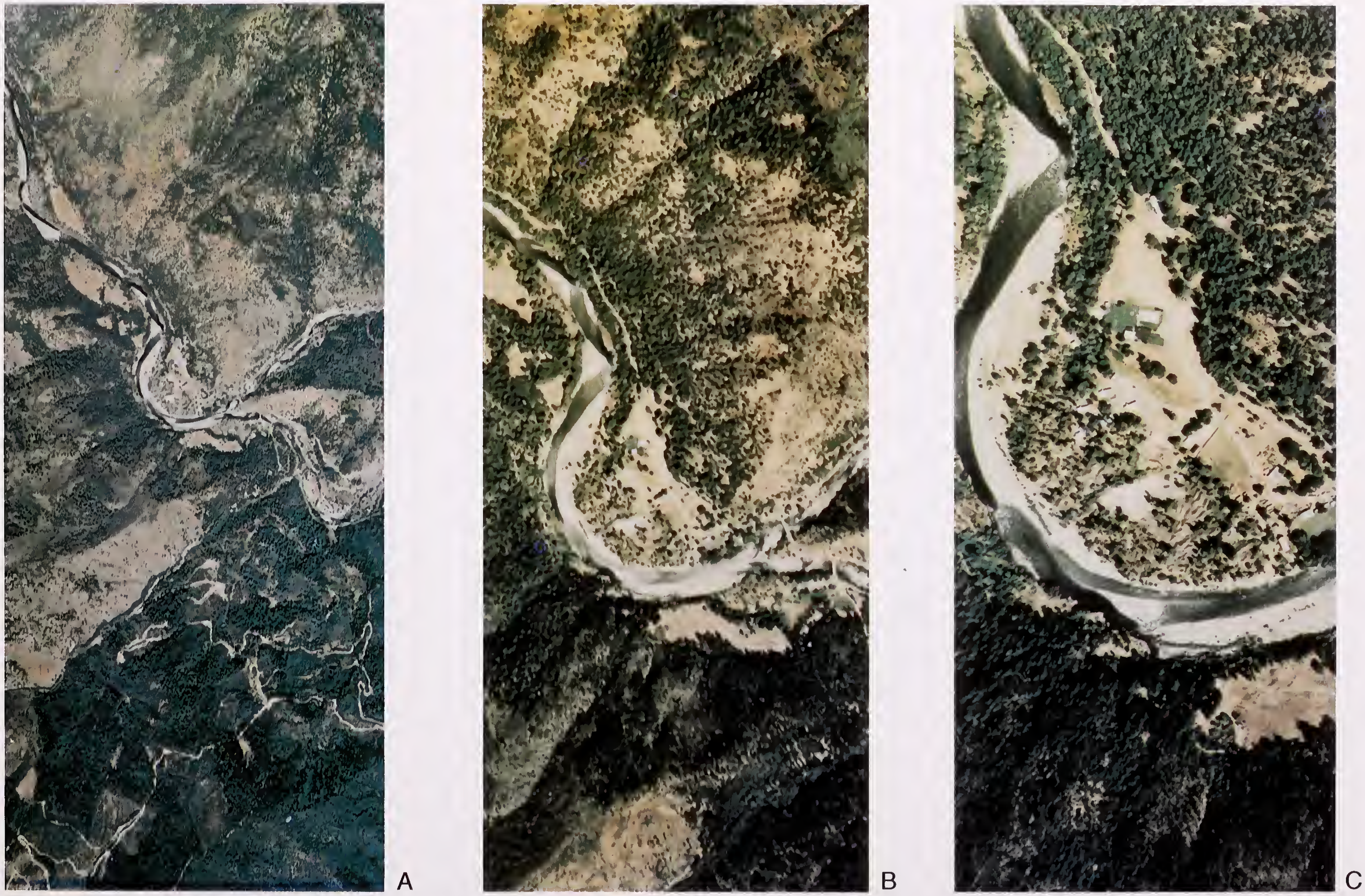
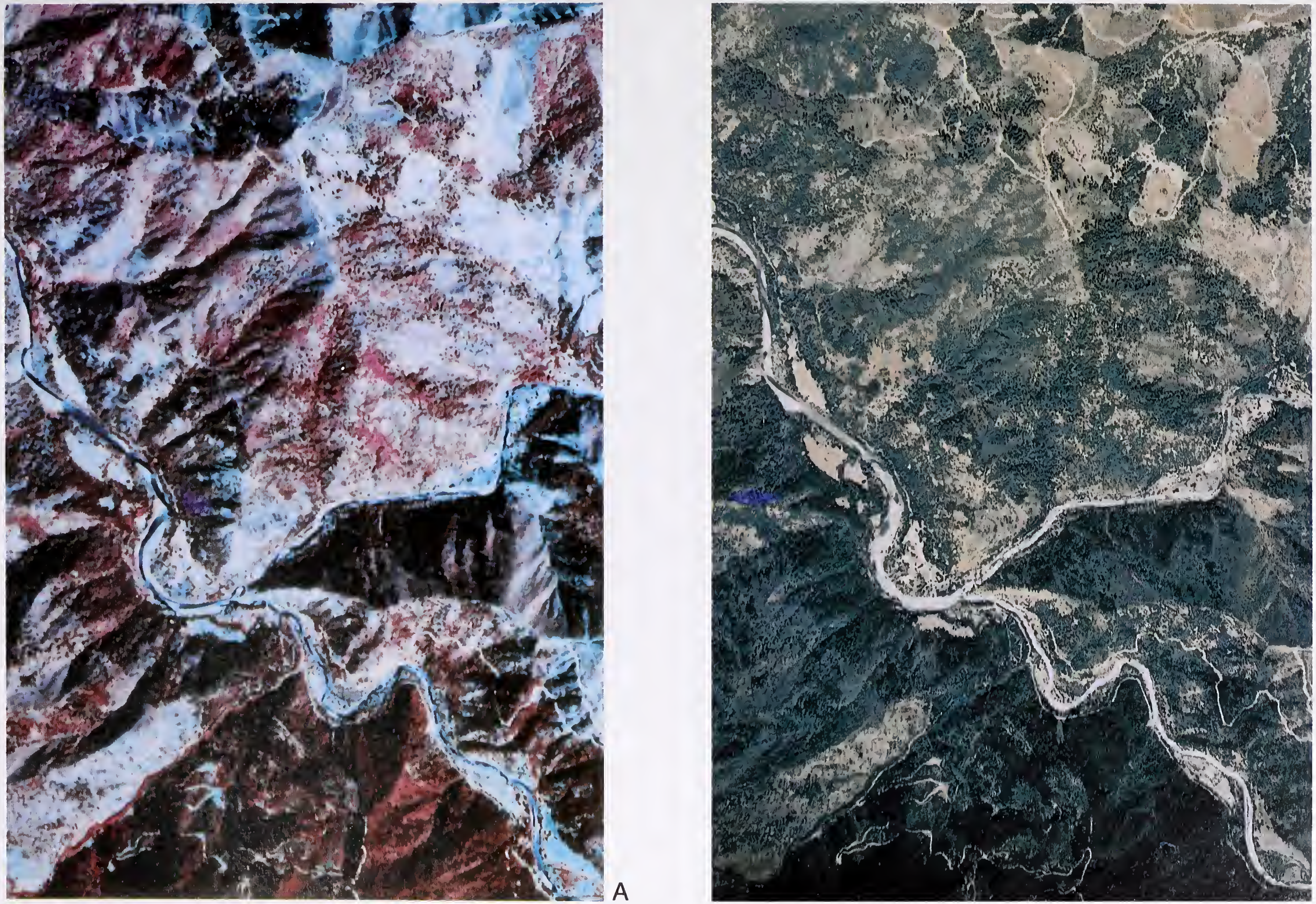


Figure 5—Illustration of scale and ground resolution. A is at 1:38,000, B is at 1:15,000, and C is at 1:6,000.



A

B

Figure 6—Illustration of film type and ground resolution. A is color infrared (CIR) and B is natural color. Photos were printed to approximately the same scale.

zones). An example of a multiple-attributed polygon coverage could be a State boundary that has segments coincident with the Enchanted Forest boundary, a compartment boundary, and a timber stand boundary. In this case, one line segment will have a nominal attribute for each coverage it appears on, may or may not be homogeneous, and must be the same line segment on all coverages.

When evaluating polygon coverages, you should distinguish between polygon boundary definition and internal variation (heterogeneity) of polygon attributes. Polygon content is either homogeneous (pure) or heterogeneous (impure). The only kinds of attributes that one may safely consider homogeneous for the discriminating attribute are political designations (such as national forest, county, and State) and perhaps water bodies. For polygons representing natural variation, homogeneity and heterogeneity are always scale dependent. All areas within a polygon on the Enchanted Forest would be homogeneous where political designation is concerned, but may be heterogeneous for vegetation type. Because many kinds of procedures (such as polygon overlay) assume that areas within polygons are homogeneous, we must know the conditions under which the homogeneity assumption is valid.

Summary

We have tried to introduce a broad if not comprehensive picture of natural resource information to be considered when evaluating existing resource information. These aspects relate to location (where), content (what), resolution (detail), and model (system) (Berry 1989), and perhaps to the age of the information (when). We recognize that there are no absolute standards of suitability and quality that apply to natural resource information. It seems highly unlikely that a universal value for any one aspect of the resource will fit all needs, and the costs of obtaining such a universal value is probably exorbitant. Whether we use the information collected in an earlier survey or inventory, or whether we collect new information depends on the value associated with particular resources at the time. Evaluation of natural resource information is possible only in relation to its intended applications and cannot possibly anticipate all future applications. Some care should be given to applying an appropriate level of measurement, not just the latest and greatest.

A careful reading of the levels and scales of measurement applied to particular mapping or GIS projects is suggested. Maps continue to have value to planners and to a broad range of the public. The accuracy and timeliness of these products can be improved by techniques and processes we have suggested. Geological Information Systems and computer applications of spatial and temporal information is still evolving. It is easy to look back and see problems that resulted from earlier applications, but the rapid change virtually guarantees that some products will be rapidly supplanted. It seems very difficult to maintain a realistic hold on what can possibly be done in the near future and what is actually needed to accomplish today's assignment. We hope this chapter has given some guidance in the process of selecting an appropriate scale, timeframe, and analytical procedure for GIS projects.

Chapter 5: Evaluating Data Suitability and Quality

Now that you, the Enchanted Forest manager, have completed an INA (chapter 2) and located existing information (chapter 3), you must evaluate the suitability and quality of the data (see figures 7 and 8). If suitable data of adequate quality are available, you must decide whether they need to be converted or updated, and you should do a benefit/cost analysis to determine whether it would be more cost-effective to gather entirely new data instead of using existing data. Finally, you should take several specific considerations into account in evaluating nonspatial data, spatial data, and remote sensing imagery.

Data Suitability

There is a difference between evaluating data suitability and evaluating data quality. Evaluation of suitability focuses on what the data purport to represent, while evaluation of quality tests to see if the data meet the purported specifications. We evaluate data suitability first because it is generally more obvious if data are not suitable for the proposed uses. There is no point in worrying about quality of classification accuracy, for example, for a set of irrelevant classes.

Checking the suitability of existing information for inclusion into an integrated resource data base can be a cumbersome and time-consuming process. We may need to consider the applicability of the data to many potential uses. Although our INA's identify information requirements and describe procedures for producing information products, we rarely specify cartographic models completely. While we may realize that different products may require similar thematic content, we may fail to consider thoroughly the issues related to scale of analysis, data resolution, and error propagation in GIS operations.

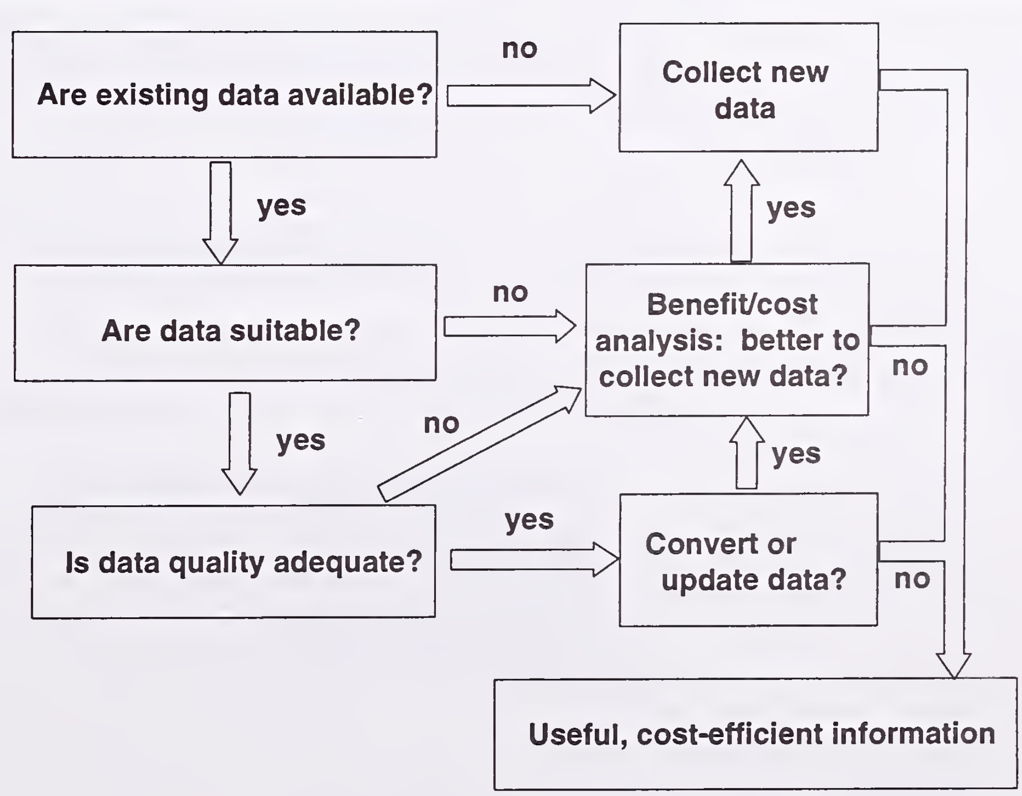


Figure 7—Flowchart for evaluating information needs.

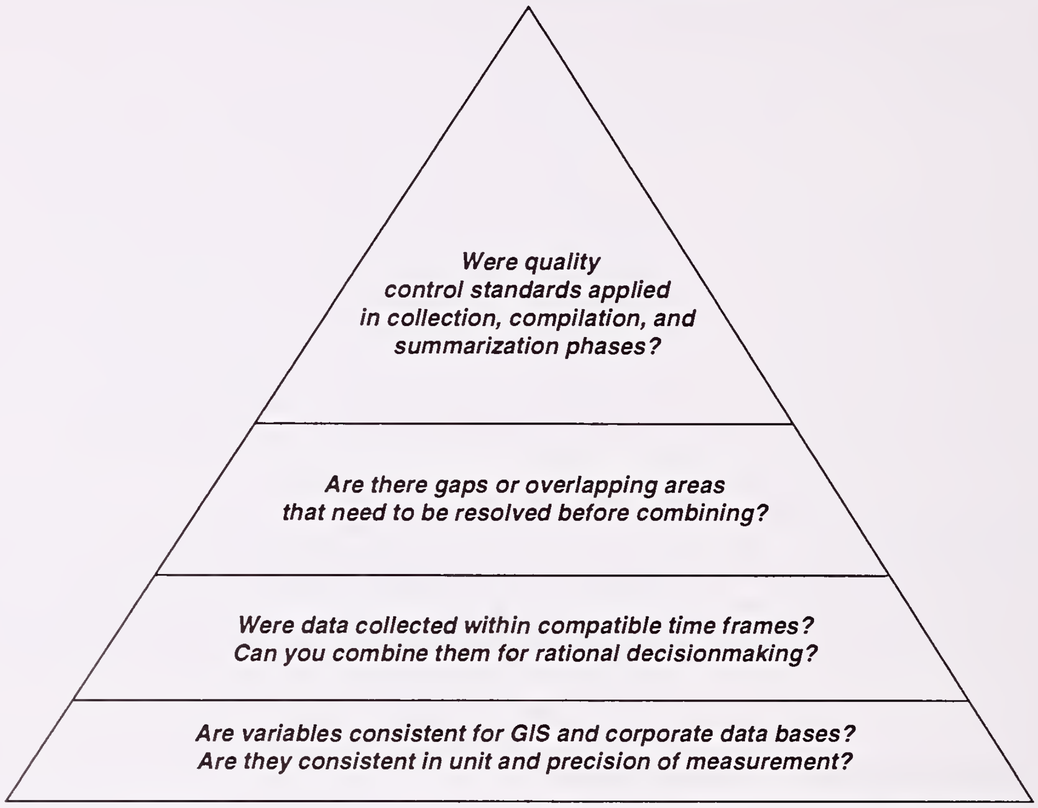


Figure 8—Data suitability evaluation pyramid.

When checking data suitability, use an approach that concentrates on the most obvious and limiting aspects of the data first. Whenever practical, begin the evaluation of data suitability with an evaluation of the primary data source. Intermediate products (such as a map) may have undergone transformations or conversions that mask inadequacies. For example, illicit edge-matching may hide gross differences in interpretations between adjacent land units; and relabeling may hide inconsistencies in the meaning of class attributes. Edge-matching should correct only for minor misalignments at map edges. It should not correct for significant differences in interpretations along map edges.

A general rule is: if the primary source is unsuitable, so are subsequent products. If the primary source seems suitable, it is still necessary to ascertain how the derived products were produced and the original intended use.

In evaluating spatial data suitability, you should consider the data's (1) thematic content, (2) resolution (level of detail), and (3) location (geographic position).

Suitability of spatial data depends on:

- Thematic content
- Resolution (detail)
- Geographic location

(Based on Berry 1989.)

Thematic Content

Begin evaluating data by looking at the thematic content. The most obvious and significant inadequacies are likely to be reflected in the thematic content. Consider measurement scales, classification systems, extent of coverage, and age of the data.

Measurement scale—To evaluate measurement scale, consider first whether the scale is appropriate for the variable being considered. For example, is an ordinal scaling applicable to this variable? Second, consider whether the measurement units and standards are consistent. For ordinal data, you should understand the criteria used to assign feature levels in the scale; terms like "high," "medium," and "low" must have a clear definition.

Classification systems—Classification systems should draw clear distinctions between groups, and if hierarchical, the user should know which variable takes precedence in defining subgroupings or delineating unit boundaries. When evaluating maps or overlays of such themes as vegetation types, soil types, or land use, consider whether the classification system is relevant to present needs. Then evaluate the accuracy of the information. Sometimes it is possible to translate (cross-walk) between two different classification systems; however, the amount of error likely to occur in such translations should be considered.

Possible relationships between categories in two classification systems include one-to-one, many-to-one, and one-to-many. One-to-one might be creek to stream, where "creek" in one classification of hydrologic features corresponds to the same features as "stream" in the second. An example of a many-to-one relationship might be lake and pond to water body, where "lakes" and "ponds" in the first classification are lumped together as "water body" in the second. An example of a one-to-many relationship might be spring to spring, seep, hot spring, and vernal pool. Classes may also overlap or gaps may exist between them. For example, there may be five levels of running water in one set and three in another.

Extent of coverage—You should also determine whether there are significant gaps in coverage in the area of interest. Has all of the area been mapped or inventoried? Are there areas for which no coverage is available? If there are gaps and if information is needed for these areas, then you must collect data for that location.

Age—The age of the data is an important thematic consideration for attributes that are likely to change over time. We may expect some vegetation information to change seasonally. Other information, such as stand structure, may not change for decades. And information such as landform may not be expected to change within a person's lifetime.

Resolution

Determining if the data source resolution (level of detail) is suitable is difficult. Resolution requirements depend on the degree of variability in attributing values that is tolerable and on the spatial frequency (how fast values change in the spatial domain). Spatial frequency may be thought of as roughness of a surface of a variable of interest. For example, on a perfectly flat plane of any extent, knowing the x, y, and z coordinates of three widely distributed points allows one to predict the elevation at any x, y location reliably. As the surface becomes bumpy, the density of sample points one needs for reliably predicting elevations increases

rapidly. Unfortunately, where there is national variability, one seldom knows the spatial frequency of the variable of interest. One must determine optimum sampling densities iteratively.

One needs to distinguish between categorical and spatial resolution, and to do so separately. Consider slope maps, for example. One can generate slope maps with 20 percent class intervals (categorical resolution) at a wide variety of spatial resolutions. Conversely, one can produce slope maps at any spatial resolution with a variety of different class intervals.

To determine the appropriate spatial resolution for a proposed application, specify the smallest land area you need to consider. Then specify what it is that you wish to know about those units. Keep in mind that one may need different levels of resolution for detection, identification, and analysis. For example, suppose that we have specified that we need vegetation information on areas 10 acres (4.047 ha) in size or greater. Do we mean that we want to detect differences between stands that are 10 acres in size or greater, do we want to identify the content of these stands, or do we want to analyze the vegetation distribution within them?

These three activities—detection, identification, and analysis—require significantly different data densities. The interpreter's rule of thumb (based on the Nyquist sampling theorem) states that one data element is sufficient for detection, provided that the image-to-background contrast is high. The threshold for identification is about 9 or 10 data elements, and the threshold for analyzing within units is about 100 data elements.

Spatial resolution also influences our ability to make generalizations. For example, when we overlay a soils map where the mapping units are on the order of 100 acres (40.47 ha) with vegetation polygons on the order of 5 acres (2 ha), we may be in a position to say something about vegetation within soils types, but not to say anything about soils within vegetation types.

Location

For most strategic-level planning, positional accuracies approach the NMAS for 1:24,000 maps. Data that we are likely to use with 1:24,000 base information should approach positional accuracies consistent with 1:24,000 mapping.

Data Quality

Current literature on the quality of natural resource data suggests that there are no generally applicable or universally accepted measures of data quality. This is partly because different uses of data require significantly different characteristics, and partly because of the complexity of the natural environment. Natural resource information can be evaluated only in light of explicit knowledge of how the information will be used and how much error or uncertainty is tolerable (i.e., what the difference is in quantitative terms between what is "good enough" and what is "not good enough"). This section discusses strategies and illustrates methods for assessing data quality and for evaluating the quality of data content.

Strategic approach to evaluating data quality:

1. Assess information with a cursory visual inspection before proceed ing to quantitative evaluation.

2. Consider requirements of integrated resource inventories.

3. Distinguish between administrative and natural resource informa tion.

4. Differentiate among direct measurements, interpolations or extrapo lations, and classified data.

5. Iterate, improve future surveys, and improve estimates based on existing data.

Strategies

Assessing data quality can be time-consuming and costly. Approach the problem strategically, focusing attention first on the most important and obvious aspects of the data. Employ more rigorous methods where there is doubt or where you must do so for analytic reasons—for example, when you need to know how uncertainty in the data impacts estimates. In evaluating data quality, follow the approach outlined in the box, above. The steps to take are discussed below.

Visual inspection—Make a cursory visual inspection of the information. Chrisman (1982) suggests that there is a continuum of rigor in evaluating data quality, ranging from deductive estimates of comparisons with internal evidence to comparison with independent source data of higher accuracy.

Inventory requirements—Because corporate data are an important goal of information management, the requirements of integrating resource information are important to consider when evaluating the quality of existing data. Integration requires that data, possibly at different times and from different areas, resource specialists, and levels in the organization, be compatible in their locational accuracy, attribute characteristics, and level of detail.

Information differences—It is helpful to distinguish between administrative and natural resource data. Administrative objects, such as campgrounds, ranger districts, and designated wilderness areas, have artificially set boundaries and are absolutely homogeneous with respect to their attributes. Natural objects, on the other hand, have indefinite boundaries that are a function of ecological processes, and may vary in position and complexity depending on the scale at which one analyzes them. Some natural boundaries are relatively definite and stable, such as the edge of a lake or beach. Others, such as boundaries between vegetation types, may be difficult to discern and may change considerably through time. Also, natural objects (such as vegetation or soil associations) are at best only relatively homogeneous with respect to their attributes.

If a land unit is absolutely homogeneous with respect to an attribute, then that attribute will also apply, without question, to any subdivision entirely within its boundaries. The same may not be said of heterogeneous units, including most

natural resource data. For example, consider a vegetation unit with 70 percent canopy closure. This attribute may not apply to any subdivision of that unit. In fact, it is impossible to determine with certainty that the attribute applies to the unit as a whole. For natural resource information, degree of homogeneity is scale dependent and always uncertain.

When checking administrative information, you need only to be sure that boundaries agree with the legal definition and conform to an accepted mapping standard. For the attributes, the only concern is that they are correct. No consideration needs to be given to the magnitude and distribution of variability within units.

Natural objects are different. Natural boundaries are not artificially set, but are interpreted. Boundary locations may vary considerably between interpreters, and locations may also vary considerably with scale of analysis.

Considerations in checking boundaries of natural resource units:

- Are the assumptions underlying the interpretation of the boundaries correct?
- Is the resolution (both spatial and categorical) of the interpretation consistent throughout the data base?
- Is the data transferred accurately from the source product to the geographic reference?

Attributes within boundaries exhibit varying degrees of heterogeneity that may or may not be randomly distributed within the units. This presents problems for operations such as overlay that further partition map units.

With administrative information, issues of locational accuracy and content can be considered separately. However, for natural resource information, locational issues and content have to be considered together. Additionally, resolution-related issues are more important and more difficult to assess. Both magnitude and spatial distribution of variability within land units are of con cern.

Categories of content information—It is important to differentiate among direct measurements, interpolations, and classified data.

Direct measurements. Direct measurements of properties may apply to point, line, or polygon features. A direct measurement at a point might be elevation. The length of the centerline of a stream segment is an example of a direct measurement of a line feature. Some direct measurements, such as population size, apply to polygons.

Interpolations—Interpolations involve estimating values of variables at unsampled locations with the area covered by the sample. The most obvious example is estimating elevations of unsampled points based on surrounding elevation samples. Similar techniques are used to interpolate depth to ground water, temperature gradients, and precipitation gradients. Burrough (1986) provides an informative chapter on spatial interpolations. Extrapolations are similar, except that they are

often predicted values at unsampled locations outside the sampled area. Predicting vegetation communities at unsampled locations based on similar site characteristics (such as elevation, slope, aspect, geology, and precipitation) is an example of an extrapolation.

Classified data—Classified data are groupings of data into categories on the basis of quantitative or qualitative similarities. Classifications may be either univariate or multivariate. A good example of spatial objects using a univariate classification is precipitation zones, which can be delineated on the basis of a single variable such as mean annual precipitation. Vegetation types are examples of multivariate classification. Vegetation type may be fairly simple, incorporating only a few variables (such as size and density of dominant species); or they may become extremely complex, incorporating a wide variety of data about species composition and stand structure.

Future surveys—It is important to use existing data to improve future surveys and estimates. Simply computing a quantitative measure of accuracy—such as the proportion of a map correctly classified—is not in itself very useful. Very low accuracy may be cause to reject an existing data base and collect new data. This, however, does not ensure that the new data will be any better. Complete data quality assessment tries to explain the sources of errors encountered and suggests ways to overcome them. You should use the knowledge about error distribution (both spatial and nonspatial) that you gain during data quality assessment to diagnose problems associated with sample, design, the interpretation process, and the mapping process. You should also use this knowledge to improve estimates based on the existing data.

To illustrate strategies for assessing data quality, consider a hypothetical project from the Enchanted Forest. Wildlife biologists are planning a $350,000 program to enhance waterfowl habitat in a 40,000-acre (16,190-ha) marshland. The project will include construction of water catchments and diversion structures. The terrain is virtually flat—relief over the entire project is only 40 feet (12 m). The forest engineer recommends using digital elevation data to create a 2-foot (0.6-m) contour-interval map for project planning and a 1-foot (0.3-m) contour-interval map for structural designs. Another government agency has, on file, a 2-foot contour-interval map of the area. The date and source of the map are unknown, but the source is at least 12 years old. There is no record of the photogrammetric control survey data or the method used to construct the map. Because collecting new photogrammetric survey data would cost at least $20,000, wildlife biologists favor using the existing map.

An estimate of map quality suggests that the map is not a reliable source of information for this project. Because there is no source for the control data, it would be impossible to relate the map to the ground reliably or to conduct a reliable accuracy assessment of the map. In addition, because there is no knowledge of the density used in the collection of elevation data, the spatial resolution of contours generated is indeterminable. Deductive reasoning alone is enough to determine that this data source is not adequate for this project.

The forest engineer contracts with a professional photogrammetrist to target and fly suitable aerial photography, obtain photogrammetric control data, and create a high-density DEM covering the project area. The DEM will be used to make the 1- and

2-foot contour maps needed. Upon completion of the mapping project, the question arises as to whether the contractor's product meets specifications. At this point, the lineage of the elevation data are well documented, but deductive reasoning alone is not enough to judge the adequacy of the data. So the engineer may request that level line survey control points be established to check the accuracy of the DEM. The inspector uses these control points to check the elevations from the contractor's product. This is an example of a comparison with internal evidence. If all control points are within acceptable margins of error for the map, the engineer may consider the product adequate. Internal evidence comparison may prove to be incorrect or meaningless with respect to external reality if the basic assumptions underlying the project specifications are erroneous.

A tiered approach to quality assessment may be applied to any kind of information. When dealing with geographically referenced information, consider not only the size, but also the spatial distribution of errors. Some parts of the data may meet standards, whereas other parts may not. A good place to start is by plotting a map of errors found in the data. Often, visual inspection is enough to determine whether there is a spatial pattern in the map errors.

Methods

In the past, when individual resource specialists constructed maps for their own use and recordkeeping, locational accuracy of natural resource maps may not have been critical, as long as users could identify map units in the field or on aerial photos. However, integrated resource inventories and GIS applications require more carefully controlled spatial relationships between different information sources. Natural resource maps should approach the accuracy standards of 40 feet (12 m) implied by 1:24,000 mapping.

Knowing the methods used to transfer data from aerial photographs to maps and the characteristics of the base map used provides a good indication of the size of error that one may expect. Ocular (eyeball) transfer of map units from photographs to topographic maps is unreliable and typically results in average errors of several hundred feet (tens of meters). For delineating boundaries that are congruent with well-defined topographic features, accuracies may be better. However, it is very difficult to estimate positions on long, steep side slopes and in gentle, undifferentiated terrain accurately. Mapping done by monoscopic or stereoscopic transfer scopes is also suspect, especially in areas of high relief.

Mapping done using orthophotographs or geocoded SPOT imagery is likely to be adequate for much natural resource mapping if the orthoproducts are well constructed. Even so, look at such maps carefully, because accuracies depend on the quality of the DEM and aerotriangulation used.

The characteristics of the base maps are also important to consider. If, for example, the base map was derived by enlarging a smaller scale map, it is less likely that the map product will be compatible with 1:24,000 maps. Moreover, many published USGS maps do not meet NMAS, nor do many updates of existing maps.

Rigorous assessment of the locational accuracy of topographic and planimetric maps is and should remain within the domain of qualified land surveyors and photogrammetrists. However, when evaluating the quality of existing natural resource maps, it is helpful to have methods for getting at least approximate measures of their locational accuracy. Such methods include visual inspection, manual overlays, and the use of root of the mean square.

Visual inspection—Careful visual inspection may be all that is needed to provide a sufficient assessment of the quality of a map product. When the map contains patterns that are readily apparent on aerial photographs (vegetation boundaries or geologic features), it is useful to lay the map over an orthoquad. Because displacements of 0.10 inch (0.25 cm) at the scale of 1:24,000 translate to 200 feet (61 m) of error, it is easy to detect gross mapping errors by this method. Remember that this is not definitive, because the accuracy of the orthophoto may also be suspect. In the future, we will register existing digital map data to digital terrain models. Then we will be able to generate overlays, registered accurately to the source aerial photos, to determine how well the data were transferred from the photos to the map base.

Manual overlays—Another quick method is to manually overlay map layers that contain lines intended to be congruent. For example, one might overlay management compartment boundaries with watershed boundaries to determine how much displacement there is between the maps along ridgelines. Similarly, one may compare adjacent maps to see how well corresponding lines meet at map edges.

Edge-matching maps and bringing lines shared between several map layers into coincidence are useful procedures, but should not be abused. Consider, for example, a line that is supposed to continue across adjacent maps but is displaced by 300 feet (90 m) at the border. Simply forcing the lines to meet hides the probability of a high degree of displacement between other features on the maps. This will undoubtedly cause problems in the future. Similarly, forcing coincident lines into agreement between map layers hides the fact that other lines on the layers are not in their correct relative locations. Use edge-matching and the alignment of coincident lines only to align objects (lines or points) that fall within the mapping tolerance on the original layers. When such displacements are consistently greater than the mapping tolerances, remap the data.

Root of the mean square error—It is often desirable to have a quantitative measure of locational accuracy, particularly when the data base is large and it is useful to document the locational accuracy of the data. The root of the mean square error (RMSE) is widely used to compute a measure of locational accuracy. Veregin (1989) discusses the use of this measure and others in detail. The RMSE compares the positions of a set of sample points on a map to the positions from a source considered to be of higher accuracy. The source may be an orthoquad, points located photogrammetrically, or points surveyed on the ground by conventional means or by GPS's.

To use the RMSE, randomly locate a set of well-defined points in both sources. For some types of maps (such as vegetation maps) it may be difficult to select readily

identifiable points randomly. In this case, identify a large number of suitable points and sample randomly from that population. For a measure of the RMSE ordinate

$$s_x = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\delta_{x_i})^2} \qquad \text{(a)}$$

where n = number of points, and δx_i is the displacement of the ith point in the x direction.

If it is crucial to know the accuracy of a map for reasons other than assessing the quality of data for inclusion in an integrated resource data base, consult qualified land surveyors or photogrammetrists. Resource specialists are better judges of other aspects of data quality.

Quality of Content

To evaluate quality of content, the quality of direct measurements and interpretations and of class data must be considered. A comparison matrix provides criteria for evaluating content quality.

Direct measurements and interpolations—The quality of direct measurements should be examined by looking to source documents or field instructions for collecting, editing, and archiving data. Evaluating the quality of these source data should be simple. Interpolated data, say for elevation, should be evaluated as to the type of interpolation applied and the required accuracy of current projects. Quite often, cubic splines or other traditional mathematical techniques are quite sufficient for natural resource needs.

Class data—Much current natural resource information exists in the form of classified or thematic maps. We often enter such information into a GIS as polygon layers. Examples include vegetation maps, soils maps, geology maps, and precipitation zone maps. Evaluation of the data can be divided into two general areas of consideration, suitability and quality. They can be best evaluated in a hierarchical approach, beginning with the suitability and then examining quality characteristics of the data (see box on page 59).

Evaluation criteria: the comparison matrix[1]—A number of evaluation methods for classification accuracy have been proposed and employed since the beginning of the remotely sensed digital (pixel) era. All are contingency tables of one sort or another, and analysis of the information they contain has gradually evolved. The kappa statistic ($\hat{\kappa}$) has gained acceptance for testing classification data; it is widely used in statistics for testing classification data and for assessment of contingency data. The following overview of some of these methods is summarized in detail by Veregin (1989).

[1] This matrix has also been called a confusion matrix. We prefer to think of it as a comparison matrix.

Evaluation of suitability and quality classifications:

Suitability issues:

- Are data categories meaningful in relation to present resource management issues?
- Are the criteria used in defining the classes known?
- Are the defining criteria relevant?

Accuracy/quality issues:

- Is the mapping repeatable? Can similarly qualified indi viduals get similar results when using the classification system?
- What is the overall classification accuracy of the map?
- How is classification error distributed among the categories?
- How is classification error distributed spatially?
- How does the uncertainty known to be in the data affect the analysis?

The comparison matrix (**C**) is a two-way table showing classification by remotely sensed techniques against a reference, or technique of higher accuracy. It is a square matrix with a row (*i*) and a column (*j*) for each class in the classification system. The number of classes is [*k*], and c_{ij} represents the number of sample units assigned to class *i* that actually belong to reference class *j* (Veregin 1989).

Table 5 illustrates the format of a comparison matrix. For the Enchanted Forest, assume that we have eight categories of land cover. We classify the forest on aerial photographs and then check the classes on the ground. The sample design for map accuracy assessment may follow a variety of strategies (Lund 1987). Use a spatially well-distributed sample, intense enough to provide a good statistical representation of all the map classes. The samples may or may not reflect the proportional representation of each class in the population (the data base).

Classification accuracy—Because there are eight classes in this system, $k = 8$. The figure 18 in the matrix cell at the intersection of sample (aerial photo) class 3c and reference (field observation) class 2, designated as $c_{5,2}$ (row 5, column 2), shows that two sample units assigned to class 3c actually belong to reference class 2. The figure 16 at $c_{5,5}$ shows that 16 sample units are correctly assigned to reference class 3c. All correct class assignments fall along the main diagonal of the matrix (from $c_{1,1}$ to $c_{k,k}$). The addition of several other rows and columns to the cross tabulation matrix complete the comparison matrix.

In table 5, the column sums (t_j) represent the number of sample units assigned to the reference class *j*. The row sums (m_i) represent the number of sample units assigned to the sample class *i*. The number of units in the sample (n) is found by summing all of the elements in the matrix or by summing the row or column sums. In this example, $n = 500$.

Table 5—Comparison matrix.

	Reference classes								m_i	$\hat{p}_i$	a_i
Sample Classes		1	2	3a	3b	3c	4a	4b	4c		
	1	52	2	4	0	0	2	0	0	60	.120	.103
	2	31	58	8	0	1	0	0	0	97	.194	.172
	3a	11	8	11	1	1	5	1	0	38	.076	.090
	3b	0	16	4	7	10	0	0	2	39	.078	.126
	3c	1	18	0	4	16	0	0	1	40	.080	.043
	4a	0	7	26	11	2	29	9	5	89	.178	.163
	4b	0	1	0	9	3	10	53	21	97	.194	.089
	4c	0	0	0	1	3	0	13	23	40	.080	.214
t_j		95	110	53	33	35	46	76	52	500	—	—

A comparison matrix may be used to assess both the classification accuracy of sample data and to get estimates of overall map accuracy. The earliest measure of classification accuracy used was the proportion correctly classified (PCC). The kappa statistic $\hat{\kappa}$ (Veregin 1989, Chrisman 1982) has statistical properties that make it preferable in most respects to older ad hoc measures. Still, there may be utility in computing various accuracy estimators and comparing them against this standard. Certainly, the matrix itself needs to be examined carefully even with $\hat{\kappa}$. We will approach accuracy assessment from a historic development of the measures of classification accuracy.

The sample PCC is simply the proportion of the sample units assigned to the correct class, the sum of the elements on the diagonal of the matrix divided by the total.

$$\hat{p} = \frac{1}{n}\sum_{i=1}^{k} c_{ii} \qquad \text{(b)}$$

where n is the total number of pixels and c_{ii} is the ith element from the diagonal. For the example from the Enchanted Forest shown in table 5, the diagonal sums to 249, and

$$\hat{p} = 249/500 = .498$$

Unless samples represent each class proportional to the area it occupies on the map, the $\hat{p}$ will not provide a meaningful estimate of the classification accuracy of the entire map. A weighted PCC ($\hat{p}_w$) accounts for the area each class occupies on the map. To compute $\hat{p}_w$ calculate a sample PCC for each class, and then weight each sample class PCC by the relative area of the class (a_i) as measured from a map. Then compute the weighted PCC for the map by summing the weighted class PCC's.

$$\hat{p}_w = \sum_{i=1}^{k} \hat{p}_i a_1 \tag{c}$$

For the Enchanted Forest example, values for $\hat{p}_i$ and a_i have been added to table 5 such that

$$\hat{p}_w = (.056 + .091 + .019 + .027 + .020 + .103 + .062 + .095) = .473$$

One objection to PCC is that it inflates classification accuracy. It fails to consider that a random assignment of classes to map units always results in a positive map accuracy (Chrisman 1982).

The kappa statistic $\hat{\kappa}$ adjusts for the correct assignments expected from a random assignment of classes (Veregin 1989, Chrisman 1982).

$$\hat{\kappa} = \frac{\hat{p} - \theta}{1 - \theta} \tag{d}$$

and

$$\theta = \frac{1}{n^2} \sum_{i=1}^{k} m_i t_i \tag{e}$$

The values of kappa lie between –1 and 1. When there are no correct assignments, $\hat{\kappa} = -1$, and when all are correct, $\hat{\kappa} = 1$. If classification is equivalent to what would be expected if class assignments were random, $\hat{\kappa} = 0$. A value of 0.5 shows that the classifier avoided 50 percent of the errors expected from a random classifier (Chrisman 1982). For the Enchanted Forest example, the sample

$$\hat{\kappa} = .498 - .138/(1 - .138) = .418.$$

As with the PCC, we may want to compute a weighted kappa ($\hat{\kappa}_w$). This requires computing $\hat{\kappa}$ for each class. Sample $\hat{\kappa}_1$ for class i ($\hat{\kappa}$) is

$$\hat{\kappa}_i = \frac{nc_{ii} - m_i t_j =_i}{nm_j - m_i t_j =_i} \tag{f}$$

Using the weighted PCC ($\hat{p}_w$) and the weighted class kappas ($\hat{\kappa}_w$), we compute $\hat{\kappa}_w$ as follows:

$$\hat{\kappa}_w = \frac{\hat{p}_w - \theta_w}{1 - \theta_w}; \ \theta = \sum_{i=1}^{k} \hat{\kappa}_i a_i \qquad \text{(g)}$$

For the Enchanted Forest example,

$$\hat{\kappa} = .473 - .129 \,/\, 1 - .129 = .395.$$

This can be interpreted to mean that we may expect the map to avoid about 40 percent of the errors made by a random classifier.

Conversion and Update Issues

When preparing to incorporate existing data into an inventory or corporate data base, you may find that information has already been collected for the area in which you wish to work. If the older data fit the standards of the data you wish to incorporate, you should consider using them and updating them, if their value is substantial. Even if the older information was collected for other purposes, you should still consider converting it to meet your needs. For example, if data were collected in a special wildlife canvass for a forest several years ago, then you as timber manager may be inclined to ignore the data and simply collect your own, because the wildlife data were not collected for timber purposes. But let's say that they did include tree diameter and height information on point-sampled plots 4 years ago. Then you should consider using these data by converting and updating them.

Is there a value to this information for the upcoming timber inventory? Almost certainly. Species occurrence frequency, diameter distribution, and volume distributions might be calculable from the wildlife inventory. Several possibilities for applying this information to the current timber inventory can be considered. Some of these issues will be discussed in more detail later; however, the main point is that combining information has become a common statistical procedure. Simple processes such as computing a required sample size and then using the data collected to compute a composite estimator probably can be accomplished by many computer-literate foresters with very little chance of error. At a more complex level, a complete Bayes analysis for timber would almost certainly involve the efforts of a specialized statistician (Bayesian). Generalized programs now exist that easily compute distributions and their complete convolution, allowing for an extremely sophisticated combination of data. See articles by Lund and Schreuder (1980), Lund (1986a, pages 39–41), and Lund (1986b) for specific questions to ask and documentation to seek.

Questions to ask before using data for updating and conversion:

- How do existing data relate to data already in corporate data bases and GIS's? Are there any training or expertise prerequisites needed to derive useful information from combining them? Can potential users readily understand the potential for combining them?

- Can the data be used without reinterpretations? (Existing information may have been created for specific purposes and contain unfamiliar jargon.)

  - Are the variables defined and used in the way required? What method should be used to combine several sources of like information? (Summary statistics and available maps may have been developed from very different classification systems.)

  - Is updating valid for these data?

  - Are the standards the same as those required? (Existing information may vary in its reliability and utility.)

Benefit/Cost Analysis

When contemplating changing or converting data bases, it is usually wise to do a benefit/cost analysis to see if it is to your benefit to convert or update the existing information or to acquire new data to replace the old. Sometimes, you may not have a choice—your organization may dictate that new data be collected, or cost and time constraints may make use of existing information unavoidable. But where you do have a choice, conduct a benefit/cost analysis.

In a benefit/cost analysis, two lists are made—one for the benefits and costs of maintaining the existing data, and one for the benefits and costs of replacing them. Then a comparison is made between the two analyses. The alternative yielding the most benefits at the least cost is usually, but not always, chosen.

On the benefit side for retaining existing data, one can list continuity with current data bases and with past programs and decisions. On the cost side, one must consider the time and effort needed to convert existing data to the requirements of the corporate data base or for entering them into a GIS. Conversion to new coding may simply require some type of computer program. However, entering maps and overlays into a GIS may require that some maps be redrawn and then digitized. In addition, if data need updating, then some of the costs of obtaining new data may have to be incurred.

On the benefit side for collecting new data, one can list the opportunity to create data bases with more flexibility and utility than those created from old data. The Pacific Northwest Region (R-6) of the Forest Service, for example, is using digital satellite imagery to create four separate layers of vegetation information (forest type, structure, canopy closure, and size class) at a pixel level. This data, when entered into a GIS, will allow users to combine and retrieve any or all of the four layers down to 1 acre in size (Green and Congalton 1991). If existing stand maps were digitized, users could retrieve only the combined data that were entered.

On the cost side of obtaining new data, one must consider the expense of purchasing new imagery and of the time, personnel, and equipment needed to perform field checks, edits, and data entry. Lund and Thomas (1989) give some rough cost estimates for conducting resource inventories.

In the Enchanted Forest, for example, suppose that data are at least 20 years old and were less than 90 percent accurate at the time of data collection. Characteristics of the vegetation have changed since, and much of the data on vegetation have not been updated. What are the benefits and costs of using existing information, and how do they compare with those of getting new data?

The benefit of using old data is their connection to past decisions. But the old data need updating, so there would be the cost of acquiring imagery of the area and interpreting it. There would also be the cost of any field surveys needed to verify the imagery interpretation and to acquire information that may not be available from other sources. In addition, there is the expense of converting existing codes to corporate standards and then entering all of the data into a GIS.

One must compare these costs to the expense of completely remapping and inventorying the forests to meet corporate standards and for entry into a GIS. The main difference between the two alternatives seems to lie in the relative benefits and costs of mapping and inventorying the entire forest or only a part of it. Some of the costs may be the same in both cases. But differences will remain between the cost of interpretation and the expense of field data collection, which will generally be higher. Once you have done the analyses, you can decide which alternative to choose.

Nonspatial Data

Nonspatial data provide information not stored on maps, overlays, or in GIS's. Usually, they are in the form of personal knowledge of an area and published inventory reports and data bases.

Personal Knowledge

Scientifically, personal knowledge and observation are always suspect and should be relied on only when no objective information has been located. Recently, a variety of computer-intensive methods loosely termed "artificial intelligence" have emerged that sometimes capture the experience and knowledge of individuals in very controlled industrial settings, but personal knowledge still must be quantified prior to reliance on such programs. It is frequently true that personal knowledge can focus a search for clearer understanding of existing data. The pitfalls of using data from another agency or internal organization may be avoided through a few well-structured interviews. An interviewer may judge the powers of observation of an expert by asking questions that seek detailed information. Lack of clarity in details might trigger the search for verification by a second or third source. The extent of an expert's experience in a specific area provides increased confidence in a particular ancillary data set provided by him or her. The longer a person has worked in an area, the more familiar he or she should be with its terrain and resources.

Inventory Reports and Data Bases

Nonspatial information includes published reports and data bases such as those maintained by Forest Service FIA's. Other potential sources of data that might be appropriate for consideration in preparing a data base include reports prepared by the FWS, Corps of Engineers, and Agricultural Research Service, which contain important forest information in areas that do not constitute high forest value. These reports could provide important information on species mix or even diameter distribution in bottom-land forests where few Forest Service plots or inventories are sufficiently intense to provide such information. Some of these agencies or other Forest Service organizations can also provide locational information and are developing extremely valuable data bases that can be referenced either by traditional methods or by online computer connections.

Questions to ask regarding nonspatial sources of data:

- When were the survey data collected? What was the duration of the survey?
- How were the plots located?
- How were the field data collected? How were they processed into the form available to you?
- Who collected the data? Who processed them?

Age and expected shelf life—Age may be the most significant problem with many existing inventories. Data for different resources may be of different ages, so apparent or potential correlations fail to achieve their expected value. Timespans of inventory periods could also cause problems, if unknown. Most age problems can be rectified through modern statistical procedures, but if ignored, they can lead to considerable misinformation. For example, when a timber inventory was recently taken, two dates of photography were used for the estimation of area. Some photographs were less than 2 years old, but many were more than 8 years old, and there was a correlation between harvested stands and age of photography. The interpretation did not distinguish between the two sets of photographs, resulting in nearly 50 percent underestimation of the harvest for the area inventoried, which was not discovered until long after the inventory was published.

Data quality—Even in nonspatial inventories and reports, location can be a significant consideration. Plot relocation may have resulted from the latest GPS; or the original plot location may be related to witness trees that are long gone, and the plot has actually been relocated in a succeeding inventory without the acknowledgment of the inventory group.

Forest inventory and analysis data have been collected for several decades. These data and the inventories based on them continue to improve. Growth and change statistics from these inventories are increasingly accurate for timber. Wildlife, range, and other resource values continue to receive increased attention in FIA surveys. Major changes in the survey and an awareness of timing and significance of changes are of considerable importance in establishing a GIS data base.

Processing data can have an important effect on their subsequent use. Raw inventory data are rarely of value to second-tier users like planners or GIS assemblers. Usually, a rather significant amount of editing and transformation from field entry is necessary before the data become useful. But too much processing can hide a multitude of sins, and highly aggregated data are often of little use to someone attempting to establish a meaningful GIS data base. For example, volumes given in older inventories for individual trees may hide the use of outdated or simply inaccurate volume equations. A more accurate recomputation of volumes may be possible if heights and diameters are given.

Benefit/cost analysis—It may be possible to compare the cost of using old or inappropriate data to the cost of collecting new data. The benefits are often not so easily quantified. Care needs to be taken that traditional benefit/cost formulas are *not* used to justify preconceived notions about the value (or lack of it) of collecting new data. Appropriate levels of application of new technology are difficult to judge; a balance between applications of old and new technology is often possible and desirable.

Spatial Data

Spatial data include maps, overlays, and remote sensing imagery. Quality or error analysis is important, not only because of the impact of error on the validity of results, but also because of its effect on operational costs. Errors resulting from misregistration of spatial data or from miscoding slow down processing. Such errors, regardless of whether they have significant impacts on the results of analysis, need to be resolved before processing can continue. In addition, measurements on a variable or set of variables may be so gross that they degrade the quality of results, thus adding to project cost while corrupting the information.

Steps to take in evaluating spatial data:

1. Identify issues relevant to the accuracy of spatial data bases, including:
   - Scale of analysis
   - Positional accuracy requirements
   - Sampling design
   - Interpretation error
   - Reliance on surrogate relationships
   - Correlated data
2. Identify obvious sources of error in spatial data.
3. Identify and evaluate existing models for assessing accuracy of spatial data.
4. Evaluate available data in relation to modeling requirements.

Maps and Overlays

Suitability of existing maps and overlays may be assessed against many elements, but probably most meaningful are content accuracy, completeness of work, and positional accuracy of features. Existing information may be used as background material for a variety of purposes, but its improper use can cause problems (see table 6). Existing data may have been originally collected for a different purpose than their current use. Manipulated data may therefore imply features that actually do not exist, or may eliminate features that actually do (Lund 1985).

Table 6—Overlay mapping problems.

Principle	Problems in overlay	Common situation
Gradual changes between polygons	Sliver errors	Delineations are inconsistent.
Multiple source map scales	Sliver errors; resolution inconsistencies	Generalizations and resolutions vary.
Interrelationship of landscape attributes	Inconsistent classes; sliver errors	Factor mapping seldom recognizes interrelationships.
Integrated landscape attributes controlling ecological processes	Key units of ecological response not captured	Selected classes irrelevant to processes.

Content accuracy concerns quality of interpretation, or how well map compilation represents actual conditions. For example, timber stands must be properly classified and road classes shown must be accurate. *Completeness* concerns how well the map shows all themes of the same class; for example, all stands of the same timber must be shown. Suitability assessed against these two factors can be determined by quality-checking a sample of work. *Positional accuracy* of any data source depends upon uses for which it is needed. GIS's are useful for at least three levels of planning: strategic (national and regional planning), tactical (forest and district planning), and project planning. Data should be accurate enough for the planning level intended, but excessive accuracy wastes time, resources, and money. Data required for strategic planning do not need to be as refined as those required for project planning. In general, data suitable for the lower (project) end of the planning hierarchy will be suitable for uses at the upper (strategic) end, but *not* vice versa.

Data for GIS's are normally referenced by a plane coordinate system and not by geodetic positions. This allows convenience in using simple plane coordinate computations rather than the more complex spherical mathematics necessary in geodetic geometry. The compromise may cause inaccuracies in results derived in GIS processing, especially for large geographic areas.

Delineation of many natural phenomena assumes discrete differences, when in reality differences are transitional. Often transition zones contribute uncertainty of position many times greater than that introduced by mapping processes.

Two aspects of concern for spatial quality are *absolute accuracy of position* and *relative precision of scale*. Absolute accuracy refers to position discrepancies with respect to some overall reference frame, such as State Plane Coordinates (SPC) or the Universal Transverse Mercator system (UTM), and may be expressed as a radius of position uncertainty at some level of probability. Relative precision concerns measurement discrepancies between pairs of points without regard to reference frame position, and may be expressed as a ratio, such as percent or parts per thousand. For project design, relative precision is often more important than absolute positional accuracy; conversely, for strategic planning, positional accuracy is more important.

Developing spatial data bases involves two types of measurement error: systematic and random. *Systematic errors* are of the same magnitude and sign for each observation (for example, scale error caused by expansion of paper documents). This type of error can be compensated for and its effect minimized. *Random errors* are variable in sign and magnitude and follow the well-known normal probability distribution. They cannot be removed, but they can be modeled. The resulting data from manual digitizing and many mapping processes are examples of data with random errors.

Spatial quality must consider capabilities of data-gathering methods. One cannot achieve better accuracy than process capability. For data gathering from PBS maps, absolute accuracy normally achievable is on the order of 25 to 50 feet (8 to 15 m) for discrete, well-defined points, and 98 to 164 feet (30 to 50 m) for continuous features. Relative precision of PBS coordinates might be on the order of 1 part in 3,000.

Map scale is probably the major factor relating to spatial accuracy of data. Ground coordinates can be measured from large-scale maps with greater precision than from small-scale ones. In cartography, 0.01 inches (0.25 mm) is considered a practical, operational limit of manual plotting accuracy. For PBS maps at 1:24,000 scale, this represents 20 feet (6 m); for secondary base series (SBS) maps at 1:126,720, it represents about 105 feet (32 m).

Map scale affects accuracy of symbol representation and placement. Symbols graphically represent either physical or administrative objects. Because they are graphical, they have dimension. For example, a forest boundary is administrative, but it is graphically represented by a line of measurable width. This line takes up much more space on the map than is taken by the boundary it represents in the physical world. Delineation of any theme presents the same dilemma. Smaller map scales increase inherent absolute error of delineation.

Some map symbols (picnic areas or campgrounds) may have no correlation to actual feature size. At scale, they may be larger or smaller than features represented, thus leading to errors of position and loss of precision. And some symbols have precedence of placement over others, so that subordinate symbols may be displaced. This cartographic problem is less common at larger scales, such as 1:24,000.

Mixing data from maps of substantially different scales is poor practice and should be avoided. Data from SBS maps are automatically more than five times "coarser" than data from PBS maps, due to scale alone. In addition, there may be compatibility problems introduced due to scale-exaggerated differences in map projections. Different coordinates will be obtained for identical points measured from maps of different scale and projection.

The following four factors also affect spatial quality (the third factor causes random error, and the other three cause systematic error):

1. *Projections and coordinate systems.* Flat maps cannot perfectly represent the spherical Earth without some distortion. Map projections are designed to minimize distortion but cannot eliminate it. Plane coordinate grids used with conformal map

projections retain true angles and match ground distances within tolerable limits, so long as measurements are confined to a fairly narrow band—160 miles (257 km) for most SPC zones. Error extremes occur at zone centers, where scale is too small, and at zone edges, where scale is too large. Scale error diminishes to zero on two lines of exact scale located about 56 miles (90 km) from center of a typical SPC zone. Typically, maximum SPC relative scale error ranges from 1 in 9,500 to 1 in well over 30,000, depending on actual SPC zone width.

It is often necessary to extend coordinate systems outside zone boundaries for consistency of data bases. Beyond zone limits, scale error rapidly becomes greater than design values. For example, scale errors degrade to about 1:2,550 on some western forests that overlap halfway into the adjacent SPC zone. This problem is more severe in Alaska, where cross-zone SPC scale errors can be worse than 1:1,000.

2. *Reduction to sea level.* Maps are cast on a datum that is nominally at sea level. This shortens higher elevation distances and introduces a relative error. For an elevation of 4,000 feet (1,220 m), the error is about 1 in 5,000; for 8,000 feet (2,440 m), it is about 1 in 2,600. At higher elevations, the shortening is greater. Depending on location in the zone, this reduction tends to compensate (at zone edge) or exacerbate (at zone center) map scale error.

3. *Base map accuracy.* Forest Service PBS maps are revisions of 1:24,000 scale (1:63,360 in Alaska) USGS quadrangles. The PBS updates vary in age from brand-new to over 15 years. All were produced under the rigid NMAS, which call for 90 percent of well-defined points to be within 40 feet (12 m) of their true position (in plan) at 1:24,000 scale, and about 106 feet (32 m) at 1:63,360 scale (American Society of Civil Engineers 1978, Department of Defense 1981). It has been assumed that *all* features shown on these maps meet these standards. This is a bad assumption, not only for discrete points, but particularly for continuous features such as streams and winding logging roads, partly due to cartographic generalization, and partly due to map compilation process limitations. A more realistic estimate of accuracy might be at least twice NMAS values. NMAS provide a consistent level of known quality.

4. *Medium stability.* Dimensional stability of various media used for map documents is a common concern, but one that is easily overcome with proper processing. Stable base material such as Mylar should be used. Paper is unstable, expanding and contracting unpredictably with changes in humidity and temperature. These changes can be corrected by registering to reference points of known position and applying a six-parameter affine transformation to correct for scale differences in x and y. Transformation programs using the so-called "rubber sheet" method should be avoided, because this artificially distorts and conceals blunders. Reference points provide geodetic control necessary for connecting to a plane coordinate reference frame, and for controlling scale of nonstable media. Without some geodetic reference, it is impossible to convert data with any confidence.

Mapping is expensive and time-consuming. Existing data should be used if they meet or can economically be made to meet minimum criteria of acceptability for your project.

Age and expected shelf life—All resource information is temporary; it's a question of when it changes—daily, weekly, monthly, yearly, or every decade, century, or millennium. Shelf life, then, is relative and depends on when and how much a given event changes the entities in a data set. Entities that we assume are permanent, such as buildings, roads, lakes, streams, or mountains, are not permanent, but do have relatively long lifespans. They often change abruptly when they do change. Contour lines, for example, might be thought of as permanent information. However, events such as severe earthquakes, volcanic eruptions, landslides, new dam construction, and cuts and fills for roads change contour lines. Rosenfeld and Cooke (1982) describe the volcanic eruption of Mt. St. Helens on May 18, 1980. An earthquake of Richter magnitude 4.9 triggered a massive landslide of the whole north side summit crater. A steam explosion followed that caused overlying rocks to surge laterally in a debris flow that filled much of the Spirit Lake Basin. As the eruption continued, the south wall of the crater was removed, decreasing the height of the mountain by 1,300 feet (400 m). Soil, rocks, and blocks of ice were tossed as far as 12 miles (20 km). Some areas were completely denuded of vegetation and covered with up to 7 feet (2 m) of ash and debris. The eruption blew down 150 square miles (388 km^2) of timber and destroyed 123 buildings and everything manmade in the Toutle Valley, including bridges and roads. Mudflows engulfed entire forests and created a delta that extended over half a mile (1 km) into Swift Reservoir. NFS and FIA biologists estimated that 146,000 acres (59,000 ha) of forest land were denuded or heavily damaged. Other natural events provide massive, long-lasting change to landscapes; the Yellowstone fires destroyed an estimated 1 billion cubic feet of timber, and Hurricane Hugo damaged 4.5 million acres (1.8 million ha) of timber in South Carolina.

But contour lines and forested area are usually fairly stable. By contrast, some data change every few months or years. For example, if timber stands on the Enchanted Forest are under active timber management and harvest, then data on these stands would change almost constantly. And if a big-game animal moves from its original home range because of a disrupting influence, then a new home range is created that is related to the first. Home ranges of a given animal may change several times during its lifespan, or may change seasonally. When the animal dies, no new information is created, and the old information becomes a historical record.

Because all resource data are temporal in nature, the real issue is determining whether the data are current enough and in good enough condition for the intended use. If not, then two courses of action are possible: developing new information, or updating existing information. The choice depends on the condition of entities in the data set and the number of entities in the data set that have changed. New information should be developed if the condition of current data is inadequate or if too many entities have changed for updating to be feasible.

Data quality—Principles for evaluating spatial data quality are shown in the following box (based on Goodchild and Gopal 1989). These principles relate to data suitability for project and strategic GIS uses (Valentine 1990).

Principles for evaluating spatial data quality:

- All spatial data are of limited accuracy.
- Precision of map analysis methods by conventional means is consistent with graphical accuracy.
- Precision of computer processing exceeds data accuracy.
- Precision of computer map analysis is inconsistent with data accuracy.
- There are at present no adequate means to describe spatial accuracy of complex features.
- A measure of the uncertainty of the results of GIS processing is needed.
- Data are easily aggregated, but less easily disaggregated, in the planning hierarchy.

Digital data cannot be more accurate than their source. Most of our PBS maps were constructed to meet NMAS, which call for *well-defined, checked points* to be within 40 feet (12 m) horizontally, at 90 percent probability (American Society of Civil Engineers 1978, Department of Defense 1981). This is a practical limit for graphical mapping technology at PBS scale. Note that this is a point standard applying to well-defined, checked points only. GSC estimates that few linear features on PBS maps have an actual continuous horizontal position quality that everywhere meets this point standard. Actual position error of features or parts of features may exceed 100 feet (30.5 m). Often, the spatial limits of continuous phenomena are treated as discrete edges, mapping a distinction not truly evident in nature. Moreover, the digitizing process itself degrades accuracy, especially hand-digitizing.

Vertical position accuracy of DEM's is widely variable, ranging from a few feet to over 100 feet (30.5 m). Factors such as age, photography, and instruments and techniques used affect DEM quality.

Project planning usually depends on data with better accuracy or resolution than provided by PBS, which is more suitable for strategic decisions than project designs. Though of little or no consequence at the strategic planning level, it is clear that, in general, PBS data are not suitable for most project design purposes.

Implied accuracy of data is also a concern. Data files commonly show coordinate values to 1 foot or less. This is fictitious accuracy, because it claims to be better than the source. Disregarding concepts of significant figures and error accumulation in digitizing and GIS software also causes invalid, misleading resolution. Resulting computer printouts to more decimal places than warranted by precision of input values hide true accuracy of computed or derived data such as acreage. True accuracy and reliability of some computer processing are actually unknown. Users of GIS-based information must be aware of these limitations. The level of decisions based on a GIS analysis must be consistent with data accuracy.

Graphical maps are visible and tangible, with readily obvious accuracy limitations. Conventional methods of map analysis are consistent with these limitations, yielding reasonable results. By contrast, digital data are invisible and intangible, making them more susceptible to misuse. Wrongly implied high accuracy exacerbates problems and risks of misuse. We must beware of improperly using high-level (strategic) data for local project planning.

Cartographic quality—Cartographic quality is important in the end use of the products, especially when maps or overlays from different sources are being combined. In this case, you should determine whether the standards are the same for each map or overlay. Does one resource use a 5-acre (2-ha) minimum area and another use a 40-acre (16-ha) minimum? If so, further analyses when the maps are overlain may be misleading and erroneous.

Two methods exist to test the quality of maps or overlays: ground truthing, and checking against known features in remotely sensed data. Ground truthing is often used to test unknown features that were mapped from remotely sensed data. A random sample can be used to obtain a statistical estimate of map quality, or the sample can be biased to determine whether the map or overlay contains correct information useful for a given application.

Items commonly tested are location, polygon delineation, attributes, and direction. For location, you look to see how close the map feature represents the actual location on the ground. Features that are mislocated cause erroneous results in overlay analyses. For polygon delineation, look to see if the boundaries of the feature are correctly drawn. If the boundary is too large or too small, inaccurate area calculations result. For attributes, features are checked for correct labeling. Features that are not properly depicted lead to invalid conclusions when used in spatial analyses. Direction is also important for contour lines. Directional errors in contour lines cause valleys and ridges and hilltops and depressions to be inverted.

Resource mapping content quality—Future projects will be performed better from lessons learned through use of current data. We will gain insights on what really is critical: which errors are tolerable and which are not, what data are actually needed, and what level of accuracy is needed for specific purposes. For example, suppose the Enchanted Forest is revising its data base. The current inventory is nearly 20 years old. Is this inventory still good, or does it need updating?

The Imperial Wizard of our Enchanted Forest has decreed that all mapping and interpretative data must meet 80 percent reliability. The forest has two methods of evaluating the data to meet the 80-percent-reliability criterion: random block sampling, and random map unit evaluations. The steps for both are outlined below.

1) *Random block mapping evaluation*
   - Step 1: Number each completed map sheet (block).
   - Step 2: Randomly choose 3 to 10 sheets.
   - Step 3: Run an evaluation traverse across representative delineations of each map unit on a map sheet. Note whether delineations represent map unit concepts. Document reasons any delineations do not represent map unit concepts.

Step 4: Summarize traverses, noting whether 80 percent of delineations crossed by all traverses represent map unit concepts. Document whether the mapping does or does not meet the required reliability.

Step 5: Carry out additional sampling required, depending on the total number of map sheets and delineations not representing map unit concepts.

2) *Random map unit evaluation*

Step 1: Number all delineations of a map unit.

Step 2: Choose 10 to 20 delineations randomly.

Step 3: Run evaluation traverses across each delineation. Note whether delineations represent map unit concepts (noted by each map unit description). Document reasons for not representing a map unit concept.

Step 4: Summarize traversing, noting whether 80 percent of all random delineations represent the map unit concept. Document whether the mapping does or does not meet the required reliability.

Step 5: Carry out additional sampling required, depending on the extent and distribution of the map unit.

The Enchanted Forest chose the random block mapping evaluation first. Ten sheets (maps) were randomly chosen. An evaluation traverse was run across representative delineations of each map unit. If delineations represented the concept of the map unit, it was documented. Likewise, if delineations did not represent the map unit concept, it was also noted. Traverses were summarized, noting whether 80 percent of delineations crossed by all traverses represented map unit concepts. Whether or not the mapping met the required reliability was then documented. Additional sampling may have been required, depending on the total number of map sheets and delineations not representing map unit concepts.

The Enchanted Forest then employed the second method, the random map unit evaluation. This was done by numbering all delineations of a map unit and randomly choosing 10. Evaluation traverses were run across each, and delineations representing (or not representing) map unit concept were noted (by each map unit description). Reasons for not representing a map unit concept were documented. Traverses were summarized, noting whether 80 percent of all random delineations represented the map unit concept. Whether the mapping did or did not meet the required reliability was then documented. Additional sampling was conducted, depending on the extent and distribution of the map unit.

If the required mapping and interpretive reliability for the inventory is 80 percent, then for a random block mapping evaluation, 80 percent of all map unit delineations must fit within the map unit description or classification system. The same may be said of the random map unit evaluation—80 percent of all delineations of that map unit must fit within the map unit description. On the Enchanted Forest, the evaluation revealed that existing mapping did not meet the desired accuracy, and problem areas were noted. Results of mapping evaluations were documented in a report containing the purpose of the evaluation, the procedure used, data records, and recommendations.

Benefit/cost analysis—In our example on the Enchanted Forest, benefit/cost analysis should then be carried out to see if an entire new mapping effort needs to be made. Benefit/cost analysis of alternative methods must be conducted in light of some measure of the quality of results. For geographic data, the effective measure of quality is accuracy in relation to the intended application of the data. Because existing mapping did not satisfy accuracy requirements for the intended application on the Enchanted Forest, at a minimum some limited new mapping efforts should be made to correct the problem areas.

Remote Sensing Imagery and Satellite Classification

Aerial photography and advanced remote sensor systems provide measurement data from which we must extract data specific to the requirements of the current information product. The goal of testing the quality of existing imagery is to determine whether the characteristics of the imagery are such that the required information can be derived with the available personnel and techniques. For this discussion, we will consider a basic requirement to provide a delineation of forest cover within a multicounty inventory unit. Forest cover data theme will be used to define strata within which ground sampling plots will be located and a basis for expanding plot volume data provided to produce survey unit summa ries.

In testing the suitability of aerial photography, coverage is our first concern. It is desirable to acquire coverage of the entire area of interest with aerial photography from a single mission or contract. The use of imagery from several missions, while sometimes necessary, can result in subtle inconsistencies in the data products derived from the imagery.

Our next concern in rating the suitability of the imagery is information content, both in absolute terms and relative to our capability to extract the information. The age of the imagery is the first factor to consider in evaluating information content. It is desirable to acquire imagery in the same year as we intend to collect ground data. This is rarely possible, especially for extensive areas. The older the imagery, the more likely that the forest cover strata will contain significant errors relative to current conditions.

If we do not compensate for the age of the imagery in the survey design, we are assuming that there is no change in the area and location of forest within the survey unit between the time the imagery was acquired and time of the field survey. The greater the time between acquisition of the imagery and collection of the ground data, the less likely this assumption is to be correct.

We should also consider the date of imagery relative to the season of the year during which the phenomena can be reliably seen when assessing data utility. For example, gypsy moth defoliation can only be estimated during the period of peak defoliation. The season of image acquisition must also be considered even with stable targets such as rock outcrops. Snow may obscure outcrops in the winter, and shadows from low solar angles may mask outcrops in fall and winter.

After determining that the mission was sufficiently recent and that imagery was acquired during the appropriate season, our next task is to determine if the information classes of interest can be reliably extracted from the imagery. In this case, we

must evaluate the spectral and spatial resolution of the data relative to information requirements. Spectral and spatial resolutions often interact in defining the likelihood of detecting a specific feature. Panchromatic aerial photography covers a single broad portion of either the visible or near-IR portion of the spectrum. Because the human eye can separate more colors than shades of gray, color photography is generally more suitable than panchromatic for photointerpretation.

A similar situation exists relative to computer-assisted interpretation. Image analysis often requires measurement values in several portions of the spectrum to separate the various features present in the scene. The intensity of reflected energy is encoded on three layers of the film emulsion for both color and CIR photographs. Depending on the season of the year and the feature of interest, one of these emulsions may be more suitable than the other. Normal color film provides information on target reflectance from the blue, green, and red portions of the spectrum, whereas CIR film provides data in green, red, and near-IR. The higher IR reflectance of hardwood foliage relative to conifers helps in separating these classes using CIR film. A wide range of features of interest in natural resource activities is more easily separable on CIR than natural color aerial photography. A similar condition exists with regard to the current generation of satellite remote-sensing systems. In general, a greater number of features are separable on Landsat TM imagery relative to the French Système Probatoire d'Observation de la Terre (SPOT) satellite. In these cases, the additional spectral bands of the Landsat TM more than compensate for the lower spatial resolution of the sensor.

The suitability of remote sensor data should also be evaluated in terms of capability to extract required information from the imagery. The skill of the image analysts or photo interpreters and the capabilities of equipment available for supporting the photo interpretation or image analysis affect the suitability of specific data to support an information requirement. In this regard, both the capability to recognize the feature of interest and the capability to extract the information and provide it in a suitable format are important. If one desires to derive forest cover delineation at a scale of 1:24,000 and only a lens stereoscope is available for manual interpretation and transfer, then it would be preferable to use transparencies closely approximating the scale of the final delineation. If, however, an analytical stereoplotter is available, smaller scale imagery requiring fewer models to cover the area of interest might be a better alternative.

Age and expected shelf life—The age of aerial photography can significantly affect its utility as a data source for resource inventories and GIS modeling. The timespan over which the aerial photography of a project area was acquired can also affect the utility of the imagery. Aerial photography missions that have taken several years to complete may be identified by date of contract. If the age and variation in date of acquisition of aerial photography are not recognized and accounted for in inventory design or the GIS analysis model, they can significantly affect accuracy of results. Substantial change in ground conditions between the date the photography was acquired and the date of the inventory can reduce the utility of the imagery as the second stage of a multistage inventory or as source of "truth" for assessing the accuracy and calibrating a satellite cover classification.

The timespan required to acquire the photographs of an area specified in an individual mission or contract can cover a period from several days to several years, depending on the size of the project area and the constraints imposed on image acquisition by mission parameters. In general, aerial photo acquisition is constrained to a 4-hour period from 10 a.m. to 2 p.m. local solar time. This requirement, along with constraints on cloud-free coverage, specified overlap between frames and flight lines, and constraints on ground cover and vegetation condition, sometimes extend acquisition over several seasons. Information on the period over which the imagery was collected should appear on index maps for the mission. The title block of individual frames of aerial photography includes the date of acquisition. The shelf life of recently acquired panchromatic aerial photography stored under recommended conditions is in excess of 100 years. The flammable base material of earlier nitrate base imagery makes it difficult to store. Properly processed and stored color aerial photographic negatives and transparencies have a shelf life of approximately 70 years. Historic imagery may lack the sharpness of more recently acquired imagery. Custom processing may be necessary to obtain the maximum detail from historic imagery.

If properly maintained and archived, digital imagery has an indefinite shelf life. The availability of archival digital data, however, is dependent on the availability of both the data and the processing systems necessary to generate user products. Because of the high volume of data from satellite and airborne remote sensing systems, the data are initially stored on high-density instrument tapes rather than industry-standard 0.5-inch (1.27-cm) computer-compatible tapes. Master hardcopy materials of satellite imagery can be produced directly from the instrument tapes. Special processing systems are required to reformat the digital data for distribution. The computer-compatible tapes for a specific scene are generated when the first request for the scene is received. As the technology advances, older processing systems are abandoned, and data in instrument tape formats are no longer available for distribution. When the earlier Landsat processing systems were updated, efforts were made to transfer a representative sample of scenes to computer-compatible tapes. To be properly maintained, computer-compatible tapes should be periodically cycled, with the data rewritten on a duplicate tape.

Quality of imagery—The contract specifications delineated by the organization acquiring aerial photography define the criteria against which the quality of the imagery is measured. The specifications of the Consolidated Farm Service Agency are used in the commercial procurement of Forest Service resource photography. Other organizations have similar specifications. The contract specifications include requirements for camera orientation, deviation of aircraft flight path from specified flight lines, image density, and camera calibration. The quality of photography acquired under contract is measured against the contract specifications prior to acceptance of the imagery. In cases where a portion of the mission was rejected subject to reflight at a later date, the rejected photography will be available, but will not meet all the requirements of the contract specifications. Photography acquired for site planning and other precision photogrammetric tasks must meet stringent quality specifications. Photography acquired for specific resource management tasks, such as forest pest detection, may not adhere to the standards established for contract aerial photography.

The production of either photographic images or digital data on computer-compatible tapes from airborne or satellite remote sensing systems is an extremely complex process. A high degree of quality control must be exercised throughout the processes to produce an acceptable product. The data listing provided by Earth Observation Satellite Corporation (EOSAT) and SPOT in response to queries on the availability of data provides information for an initial assessment of data quality. Each system provides a subjective estimate of the proportion of the scene covered by clouds or cloud shadows. If alternative data sets are available, scenes with more than 10 percent cloud cover are generally rejected. Microfiche of individual images can be examined to determine the location of clouds in relation to the user's area of interest. Query listings also provide a numerical assessment of image quality. These ratings can be used to reject data sets of obviously poor quality. If a quality rating is low or the band is shown as missing, it will probably be a poor candidate for production of a digital data set.

Catalog descriptions are no guarantee that a product of acceptable quality will be produced. It is the user's responsibility to assess the quality of both digital and photographic products upon receipt. In cases where the products delivered to the user do not meet quality standards, the data provider will generally remake the product or substitute an alternate scene.

Remote sensing products should be examined to determine whether sufficient contrast is present to separate features of interest. Improper gain setting during data collection or the use of inappropriate lookup tables in product generation can result in images with low contrast. Atmospheric conditions, such as high levels of haze (which can reduce the accuracy of manual or digital image classification), can only be detected by examining the imagery. Other data quality problems include periodic dropout of data. Individual bands must also be assessed. Users are generally offered a choice of several levels of geometric corrections when ordering digital satellite data. Upon receipt of the data, the user should determine if the positional accuracy of the data are within the limits specified for the product requested.

Classification systems—Estimates of area occupied by different forest and cover types are used for certain planning assessments on national forests, primarily by the Supervisor's Office. The Forest Service analytical forest planning model is a familiar example of an analysis model designed for this purpose. Similar types of analysis models will probably be needed for many more years. Remotely sensed maps, GIS data bases, and inventory sample plots can provide the required areal estimates for each cover category. For example, the percentage of Landsat pixels classified as a cover type can be used to estimate the true area occupied by that cover type. Remotely sensed areal estimates are often treated as unbiased estimates of the true area for each cover type. However, classification errors do bias area estimates (Card 1982, Chrisman 1982, Hay 1988, Czaplewski and Catts 1990).

The first objective of this section is to present an informal method to quantify the expected magnitude of misclassification bias. With this method, you can judge the practical importance of anticipated misclassification bias in remotely sensed areal estimates relative to the accuracy required by the analysis models. If the anticipated misclassification bias is unacceptable, then areal estimates can be statistically

calibrated, using remotely sensed and reference classifications for a representative sample of plots. The second objective of this section is to introduce existing formal methods that statistically calibrate biased areal estimates.

Defining misclassification bias—Misclassification bias is closely related to the classification accuracy that was discussed above in the section on evaluation criteria and the comparison matrix. Accuracy assesses the overall difference; misclassification bias is determined from similar formulas as the difference between the standard and the remotely sensed estimate. They can be computed from the elements of the comparison matrix, column totals from the comparison matrix, and simple matrix algebra.

The magnitude of misclassification bias can be predicted with statistical estimators, some of which are similar to Equation (b) (page 98). Calibration is described correctly and in detail in Veregin (1989).

$$t_j = t_{jj} = \sum_{i=1}^{k} c_{ij} \tag{h}$$

and

$$e_{ij} = c_{ij} / t_j \tag{i}$$

The two equations provide cell values for k x k matrices, which are called **T** and **E,** respectively. **E** can be obtained by matrix algebra as $\mathbf{E} = \mathbf{T}\mathbf{C}^{-1}$. Finally, **R** is a vector of length k where r_i is the number of pixels assigned to class i on the image.

To obtain a corrected class estimate, perform the matrix multiplication $\mathbf{A} = \mathbf{E}^{-1}\mathbf{R}$; matrix **A** is the k x 1 matrix of corrected area estimates. Note that there are a number of variants on this procedure, and numerical problems with matrix inversion may need to be addressed (Veregin 1989).

Anticipating magnitude of misclassification bias—Figure 9 portrays the magnitude of misclassification bias for a wide range of classification accuracies. Figure 9 can be used to anticipate the approximate magnitude of misclassification bias for any cover type, given your expectations of prevalence of various cover types and the remote sensing specialist's expectations of classification accuracies. This process can be repeated for each cover type, and examples are given in a later portion of this section. If the anticipated misclassification bias for any cover type is unacceptable, then more formal calibration techniques should be considered. Many of these statistical techniques are given in the remainder of this section.

The classification bias can be determined from the following equation:

$$Y = H_A * X_A + [(1 - H_B)*(100 - X)] \tag{j}$$

where H_A and H_B represent the correctly classified and interpreted proportions of types A and B, respectively; and X_A represents the true percentage of cover type A. (The value of $100 - X_A$ is also X_B). Misclassification bias is the sum of the vector of $Y - X$ values for the classification.

In figure 9, remotely sensed estimate (Y) is a function of producer's classification accuracies (H_A and H_B) and prevalence of a cover type (X), as given in Equation (j). This figure can be informally used to anticipate the magnitude of misclassification bias, given approximate expectations of classification accuracy and prevalence of cover types. If the anticipated magnitude is unacceptable to you, then you should consider formal calibration methods to correct for misclassification bias.

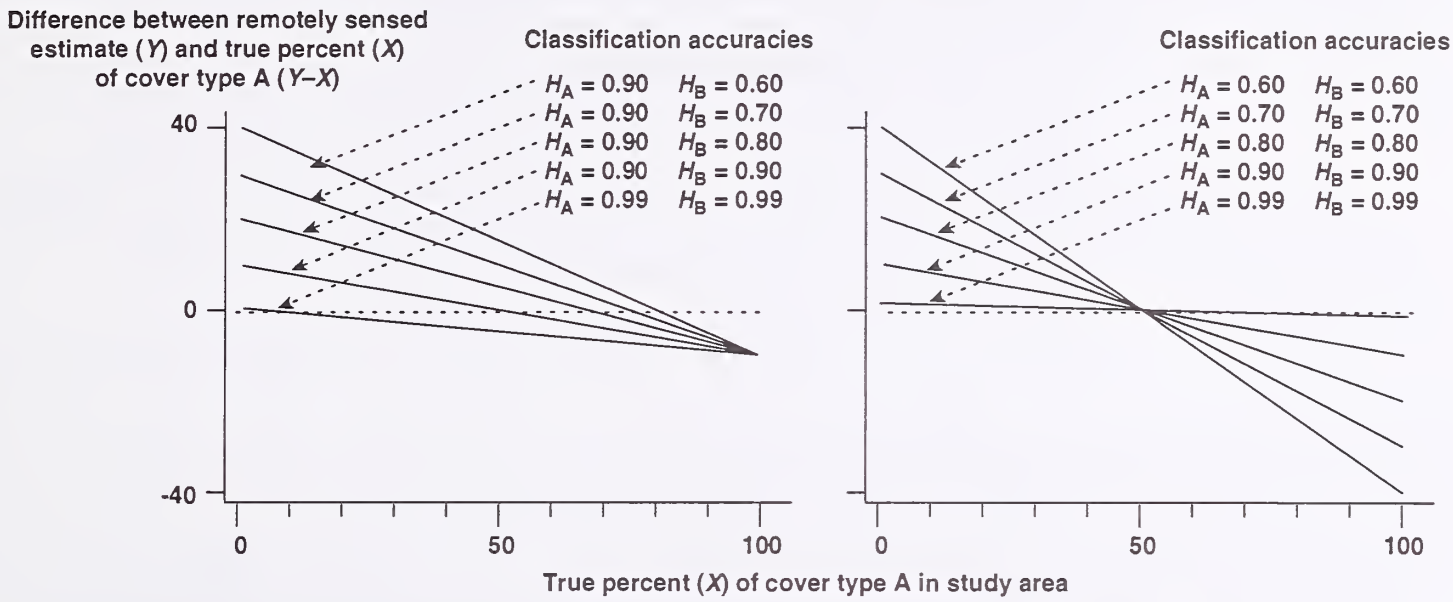


Figure 9—Misclassification bias is the difference between the remotely sensed estimate and the true percentage of a cover type.

Correcting misclassification bias—The magnitude of misclassification bias can be predicted with statistical estimators, some of which are similar to Equation (j). One cannot identify misclassified pixels with calibration. Calibration is a probabilistic technique that uses percentages of imperfectly classified pixels in various cover types to predict the true percentage of each cover type. Calibration requires misclassification probabilities that can be accurately estimated with a sufficiently large and representative sample of reference plots.

Grassia and Sundberg (1982) present a classical multivariate calibration estimator, which has been applied in remote sensing by Bauer and others (1978), Maxim and others (1981), Prisley and Smith (1987), and Hay (1988). Equation (j) is an example of a classical univariate calibration model. To produce the calibrated areal estimate, Equation (j) is solved for the true percentage (X) given the remotely sensed estimate (Y) and accurate estimates of the probabilities of omission errors (i.e., H_A and H_B).

Card (1982) and Chrisman (1982) have applied the inverse calibration estimator of Tenenbein (1972), which is an alternative to the classical calibration estimator. Unlike the classical estimator, this inverse estimator uses probabilities of commission errors (i.e., user's accuracies) rather than omission errors.

Consider a hypothetical inventory project in which remote sensing is used to map woodland vegetation on the Enchanted Forest. All pixels are classified into one of only two categories: woodland (Wd) and other cover types (Ot). Before this project is conducted, you expect that approximately 25 percent of the forest is woodland, and the remote sensing specialist expects producer's accuracy of 95 percent (0.95)

for both cover types (Wd and Ot). Equations (h) and (i) can be used to predict the magnitude of misclassification bias. In the case of Wd, $p_i = 0.25$, Ot = 0.75, and $e_{jj} = 0.95$; hence $e_{ij} = 0.05$. Application of the correction formula for these two categories can be done by hand. The inverted, transposed **E** matrix multiplied by the proportions in the two classes yields 27.5 percent for Wd and 72.5 for Ot.

You might judge that this magnitude of bias is acceptable for planning purposes, and that calibration is unnecessary. But if you judge that this magnitude is unacceptable (perhaps because woodland is critical habitat for an endangered species on your national forest), then calibration is required. Assume that you establish a random sample of 282 reference plots uniformly across the entire Enchanted Forest to estimate probabilities of various misclassification errors. A woodland map is produced for the forest from remotely sensed imagery. Twenty percent of the pixels for the entire forest are classed as Wd, based on the remote sensing image. The 282 reference plots were classed using both ground truth and remote sensing. The results are displayed in table 7 in the form of a comparison matrix.

Table 7—*Comparison matrix for two-category classification*

	Remote sensing class		
Field class	*Wd*	*Ot*	*Total*
Wd	47	5	52
Ot	6	224	230
Total	53	229	282

Use the column proportions to correct the estimate. For Wd, 47/53 (0.887) is correct, and for Ot, it is 224/229 (0.978); *note* that the complement is used in computing the corrected estimate.

$$\begin{aligned} X_{\mathrm{Wd}} &= .887(20) + (1-0.978)(80) \\ &= 17.74 + 1.74 \\ &= 19.48 \text{ percent} \end{aligned}$$

This is a correction of about half a percent. Similarly, correction of Ot proceeds as

$$\begin{aligned} X_{\mathrm{Ot}} &= .113(20) + (.978)(20) \\ &= 2.26 + 78.26 \\ &= 80.52 \text{ percent} \end{aligned}$$

The correction or adjustment is relatively simple, and the result continues to sum to 100 percent of the study area.

This example contains only two categories of land cover. However, detailed classification systems are more typical. For example, “Other cover types” (Ot) might be subdivided into the following 10 categories: woodland (Wd), riparian (Rp), pasture (Pa), barren (Ba), developed (Dv), planted pine (PP), natural pine (NP), oak pine (OP), bottom-land hardwood (BH), upland hardwood (UH), and nonstocked (NS).

Calibration using homogeneous reference plots—There are many methods for calibrating remotely sensed imagery (Veregin 1989). Some use direct counts of pixels, while others use proportions, and there are problems to be avoided when using almost any of them. Inverse and classification calibration will produce slightly different estimates (Czaplewski and Catts 1990). Based on Monte Carlo simulations, the inverse calibration estimator of Tenenbein (1972) is less biased, more precise, and less prone to numerical problems and infeasible solutions, especially for small sample sizes (i.e., 100 to 2,000 reference plots). For example, the classical estimator of Grassia and Sundberg (1982) can produce negative areal estimates, but Tenenbein's (1972) inverse estimator will always produce positive estimates (Czaplewski and Catts 1992).

All of these calibration techniques are closely related to various multistage or multiphase sampling designs, which can be more efficient than calibration if the sample size of reference plots is large. Consultation with a statistician familiar with remotely sensed data may be an important decision in selecting correction procedures appropriate for your situation.

Some remote sensing specialists recommend that misclassification bias be ignored if classification accuracy is high; however, misclassification bias will almost always occur, even when accuracy is high. During early stages of project planning, you and the remote sensing specialist should anticipate the magnitude of misclassification bias, perhaps with the informal methods given in this section. If the anticipated magnitude is unacceptable, then your project plans should include statistical methods to calibrate the final areal estimates with reference plots, perhaps with the formal techniques cited (Veregin 1989, Czaplewski and Catts 1992, Grassia and Sundberg 1982).

Calibration using heterogeneous reference plots—Reference classifications for each pixel within the reference plots might not be available, or registration error may be large relative to interpretation error. For example, reference data for agricultural surveys can be limited to areal estimates of different crop covers within large, heterogeneous reference plots; maps showing the location of each crop cover within the reference plots might not be available. Therefore, the reference classifications for each pixel within the reference plots are not available, and the calibration estimators cited above cannot be used. Here, calibration can only use the remotely sensed and reference percentages of each cover type within each reference plot (Chhikara and others 1986, Heydorn and Takacs 1986, Hung and Fuller 1987, Battese and others 1988, Chhikara and Deng 1988). Similar situations arise when registration of pixels to reference plots is difficult, and reference classifications for individual pixels cannot be obtained (Iverson and others 1989). A current example is the problem of obtaining reference classes for advanced very high resolution radiometer data. Remotely sensed estimates from Landsat scenes served as the reference data. A related situation occurs for mixed pixels that cannot be classified into unique categories with reference data; Pech and others (1986) give a calibration estimator to estimate percent vegetation cover in arid lands from mixed pixels. All these methods are regression estimators rather than the probabilistic techniques of Tenenbein (1972) and Grassia and Sundberg (1982).

Calibration based on regression methods can produce negative areal estimates. Lewis and Odell (1971), Liew (1976), and Shim (1983) propose quadratic programming techniques to avoid negative estimates, and Langley and others (1980) have applied this solution in remote sensing. Computer algorithms for quadratic programming are available in an increasing number of mathematical libraries and can be programmed in S-Plus, Gauss, and other languages; consultation with a programmer or statistician would be wise.

These calibration techniques are closely related to multistage or multiphase sampling designs, which, as previously pointed out, can be more efficient than calibration if the sample size of reference plots is large. The remotely sensed data are analogous to the first level of a multilevel design, and the reference data are analogous to the second level. However, calibration methods have been developed to use areal estimates from wall-to-wall imagery and for multivariate and nonlinear situations; calibration might be more easily applied to these more complicated estimation problems than multilevel sampling designs. You should consult with a statistician experienced in survey sampling if you are considering these more efficient, but more restrictive, techniques. The same suggestions involving misclassification bias applicable to homogeneous reference plots apply to heterogeneous reference plots as well.

Benefit/cost analysis procedures for obtaining new coverage—It is not always easy to determine whether or not to obtain new coverage of an area by using benefit/cost analyses. Complications may arise because a multistage sampling of the area is desired that requires satellite, aerial, and ground levels. Old photographic imagery may cover the area in question when combined with more recent satellite imagery, but may not reflect current conditions. There may often come a point where the analysis is more involved and takes longer than is allowed by the project. In this case, the only possibility may be an informal discussion with specialists on general costs and probable benefits.

Summary

Data suitability and quality should be evaluated separately for spatial and nonspatial data. Data suitability refers to how well information may meet your needs. When evaluating suitability, you should consider three criteria–thematic content, resolution, and location. Data quality addresses how good the information is. Methods for evaluating quality of information include visual inspection, manual overlays, and the use of root of the mean square. Methods for evaluating quality of content include direct measurements and interpretations.

Nonspatial information can be obtained from personal knowledge, inventory reports, and data bases. Spatial information consists of digital data (such as that in a GIS), maps and overlays, and remote sensing. Because the sources are different, and because each source has unique problems, each must be dealt with differently. In addition to suitability and quality of the content, one must consider the benefits and costs of obtaining new or replacement data before rejecting existing information for a particular use. You should carefully evaluate existing data in the context of its current application or need before using it in corporate data bases, a GIS, and for designing resource inventories.

We have presented some examples that demonstrate increasing complexity of analysis in several information areas. Comparison matrices demonstrate some of the classification accuracy problems in a GIS. Methods for combining estimates to improve or update inventory information are demonstrated. We mention advanced methods for obtaining updated information, including Bayesian analyses. These methods are becoming increasingly available to practitioners in the form of computer programs that guide one through an analysis. On the other hand, a simple straightforward analysis procedure that is well accepted should be given adequate consideration. The consumer of information should neither accept a black-box analysis nor worry about complex analysis or its explication for most quality and suitability issues in the use or updating of information.

The value of existing information for future projects should be recognized by both management and practitioners. Respect for past data collection is an important first step. Scientific and statistical improvements make it increasingly easy and appropriate to incorporate or use existing information. Cultivating an attitude that we can incorporate past information into current efforts may help to restore respect for the efforts of previous projects while also improving the products we need to do our job.

It is helpful to have a general idea of the potential for implementing procedures to include or use existing information. The descriptions in the following box should help you to decide whether to use existing information and to avoid the expense of fruitless applications. There are many agencies and departments within the Federal government as well as some State agencies that have information in a form that could contribute appreciably to the development of new corporate and GIS data bases. Existing information may be used as background material for inventory research, as auxiliary input for correlation studies or survey designs, to validate new research, for resource analysis, and as direct input to reports, data bases, and GIS's.

Types of existing information and ease of application:

Easy	*Moderate*	*Difficult*
Well-established methods exist and are employed. Standards exist and are adhered to by trained, certified practitioners.	Standard survey procedures with known accuracy are followed.	No standard procedures are followed, and there are no trained, certified practitioners.
A concise terminology is used.	Terminology is variable and may differ among disciplines.	Terminology is discipline-specific.
Features are static and long-lived.	Features are static and exceptions are controllable (such as road locations).	Features are dynamic, not controlled.
Regions are homogeneous.	Regions are classed from continuous data; information may be lost on recombination.	Regions are heterogeneous; processes are spatially random. Results are scale-dependent and will be used to predict future conditions.
Examples: road design, cadastral survey, site plans.	*Examples*: hydrography, transportation system maps, administrative sites.	*Examples*: economic modeling, vegetation mapping.

(Based on Calkins 1983.)

Converting Data for GIS and Corporate Data Bases

Conversion or incorporation of existing data into data bases has, in the past, required large amounts of time and effort. The bulk of this effort has been the digitizing of information that existed as line drawings or maps or other physical pieces of data. While scanner and digital video technology seems to be making progress toward automating some of the procedures for transforming hardcopy maps and overlays into digital form for storage in data bases, a basic understanding of the processes used in the past may help you decide whether data are amenable to translation or transportation into GIS's or corporate data bases.

Personal Knowledge

There are instances where first-hand, expert knowledge provides us with the most pertinent information, if it can be converted into digital form or variables. The process for incorporation is similar to any conversion of personal information. First, determine that you and the person who possesses the information are using the same terms and definitions. Though this may seem trivial, take time to ascertain that you both are speaking about the same attributes. Next, have the person relating information to you annotate the information on maps, overlays, or photos of the area. Then the information can be digitized for entry into a GIS.

Aerial Photos and Imagery

The cost of preparing and entering existing stand maps into a GIS runs about $0.02 to $0.04 per acre ($0.05 to $0.10 per ha) (Bain 1991). If new maps are first created using traditional aerial photo interpretation and field surveys, then the total costs rise to $0.06 to $0.09 cents per acre ($0.15 to $0.22 per ha) for stands averaging 15 to 20 acres (6 to 8 ha) in size.

The Pacific Northwest Region (R-6) of the Forest Service has created "stand maps" using Landsat TM data and field surveys. Because the data are already in digital form, they are suitable for use in a GIS with minimum work. Vegetation is displayed as four separate layers in the GIS: forest type, crown closure, stand structure, and stand size (diameter) class (Green and Congalton 1991). The resolution for the R-6 data is at the pixel level, or 98 to 262 feet (30 to 80 m). The costs for using the Landsat imagery range from $0.14 to $0.30 per acre ($0.35 to $0.74 per ha) (Teply 1991). While the costs of this newer technology are higher than those for traditional stand mapping, the data bases created are more versatile and flexible. Along with the increase of information comes an increase in data storage needs. The following discussion is based on Landrum and others (1991).

Table 8 lists file sizes for the digital data most commonly available within the Forest Service for an average 7.5-minute quadrangle map (Landrum and others 1991). As the numbers in table 8 demonstrate, data resolution directly determines file size for a given area of coverage. As spatial resolution (pixel size) and spectral resolution (number of bands, bits per pixel) increase, file size increases dramatically.

Using the general file sizes shown in table 8, the size of a hypothetical set of data can be estimated for a 7.5-minute quadrangle map at "medium" and "higher" resolution (where 1 meter = 3.28 feet), as shown in the box on page 87. Data storage requirements for a hypothetical 30-quad study area containing the data layers would thus be in the range of 210 to 900 MB.

Table 8—File sizes of various digital data for a Forest Service/USGS 7.5' quadrangle

Data source	Data type	Pixel size	File size (mbytes)
Forest Service cartographic primary feature file base layers	GSC vector format	N/A	0.8–1.2
Forest Service/USGS digital elevation model	16-bit raster	30m	0.4
	16-bit raster	25m	0.5
LANDSAT multispectral scanner, 4 bands	8-bit raster	80m	0.1
LANDSAT thematic mapper, 7 bands	8-bit raster	30m	1.25
	8-bit raster	25m	1.75
SPOT panchromatic, 1 band	8-bit raster	10m	1.6
Generic resource layer (derived from any of the above)	4-bit raster	10m	0.8
	4-bit raster	25m	0.12
	8-bit raster	10m	1.6
	8-bit raster	25m	0.25
	16-bit raster	10m	3.2
	16-bit raster	25m	0.5
Scanned aerial photographs			
Black and white, 1 band	8-bit raster	5m	6.4
Color/color infrared, 3 bands	8-bit raster	5m	19.2
Black and white, 1 band	8-bit raster	2m	40.0
Color/color infrared, 3 bands	8-bit raster	2m	120.0

Storage requirements for similar-size imagery:

Medium-resolution data set (25m pixels):	*MB/quad*
Primary layers from cartographic feature file	1.0
Digital elevation model @ 25m, 16-bit	0.5
Landsat thematic mapper data @ 25m, 8-bit	1.75
10 resource layers @ 25m, 8-bit	2.5
10 resource layers @ 25m, 4-bit	1.25
Storage total	7.0

High-resolution data set (10m pixels):	*MB/quad*
Primary layers from cartographic feature file	1.0
Digital elevation model @ 10m, 16-bit	3.2
SPOT panchromatic data @ 10m, 8-bit	1.6
10 resource layers @ 10m, 8-bit	16.0
10 resource layers @ 10m, 4-bit	8.0
Storage total	29.8

In order to efficiently analyze and manipulate this data, a GIS system will need large amounts of space for intermediate files, temporary files, plot files, etc. Depending on the software system being used, the amount of temporary storage space needed for analysis may be 5 to 15 times the size of the data to be analyzed.

The large amounts of storage required for a GIS mean that both online and offline data storage are required. Online storage is directly accessible by the computer and is typically read/write-capable. Hard drives and some optical disk drives are examples. Offline storage is not directly accessible by the computer and has traditionally been for archiving and transferring data. Floppies, reel tapes, and cartridge tapes are typical offline storage media. Fortunately, new technologies, especially erasable optical drives, are blurring the distinction between online and offline devices.

Having discussed GIS data requirements, we offer the following guidelines for use when considering data storage needs for projects involving digital remotely sensed data:

1. *Provide adequate online storage for the completion of the largest feasible project that is envisioned.* For example, if project data will take up 300 megabytes, and if processing will consume 5 times that much space at certain steps in the analysis, then 1.8 to 2.0 gigabytes of available online storage space would be required.

2. *Keeping the above guideline in mind, be conservative when estimating needed online storage space.* Include provisions for multiple copies of important files, system overhead, software, and uncertainty about a particular software package's storage methods, as well as space for the project data and their analysis.

3. *Provide for an efficient offline storage system that allows data sets to be archived, reloaded, analyzed, and rearchived with a minimum of trouble.* Optical drives and tape cartridges are well suited to this task.

4. *Recognize that online storage is a very versatile and valuable resource if available.* There is always a use for "excess" disk space, but lack of adequate online storage can severely limit the timely completion of a project by complicating processing steps, requiring constant attention to transfers to and from offline data storage devices, and even discouraging personnel from utilizing higher resolution data at the expense of disk space.

Maps and Overlays

Existing maps and map media can have a high value for establishing and improving GIS or corporate digital data bases. Consideration of the cost of incorporating these data will be important to the quality of the product.

In a GIS project, there are four principal ways to improve quality of the final products while reducing costs and improving timeliness. First, eliminating steps increases efficiency. Each step in the conversion process (data collection, preparation, processing, and analysis) costs money, takes time, and may degrade quality. Second, eliminating variables may help. It is difficult to measure some variables, and some are only marginally important, so the analysis may be better off without them. Third, reducing resolution is often appropriate. Using unnecessarily high resolution data can degrade results by increasing the potential for errors, and costs increase as scale and density of data increase. Working at near-optimal resolution reduces errors and costs while allowing higher resolution data to be obtained where

they are actually needed. Finally, using proven methods is best. Often, it seems quicker and less costly to shortcut professionally accepted procedures. But the apparent gains are often lost if the work needs to be redone, resulting in lost time, expensive editing, and failure of results to meet standards.

Reduce costs and improve quality by:
1. Eliminating unnecessary steps
2. Eliminating variables
3. Reducing resolution as appropriate
4. Using proven methods

Geological Information Systems are used to improve resource planning and management. A GIS operates on data—an enormous amount of it. Many data files are available to users, such as the primary base map layers digitized by the GSC. Other data files are obtained by contracting and some by purchase from other agencies and sources. But a lot of data must be obtained by inhouse digitizing, either to create original data files or to update existing ones. Digital data are expensive. It is important to convert data correctly to avoid costs of rework and also to avoid mistakes that can remain hidden for several years, buried deep in computer data bases. This section should help field units become aware of some pitfalls and mistakes that can be made in developing a GIS and help streamline GIS implementation.

***Raster or vector*—**Tablet digitizing and scanning are alternative methods for encoding maps and overlays. In tablet digitizing, the operator manually traces each line on the map overlay with the puck or cursor of the digitizer tablet. In this process, a coordinate value is transferred to the computer at preset time intervals or when the operator pushes a button on the puck. Tablet digitizers are relatively inexpensive and suitable for input of a wide range of overlays for both GIS development and update. In scan digitizing, an image of the overlay is captured by the scanner. The operator uses a combination of manual and semiautomated procedures at the workstation to extract the coordinates of the linear features from the image of the map overlay. The cost of equipment and the level of operator training required are significantly higher for scan digitizing than for tablet digitizing. Scanning is most appropriate for overlays with a high density of linear or polygon features. Soil survey maps and contour plates are examples of overlays well suited to scan digitizing. This section focuses on techniques for tablet digitizing. The principles are also appropriate to scan digitizing projects.

***Controlling ("boss") layers*—**Certain mapped features have greater validity or accuracy of position than others. For example, timber type lines often are better seen on the ground (hence better defined) than soil type lines. These features and the layers that contain them should control boundaries for features that coincide with them. Generally, features shown on cartographic layers have the highest positional accuracy because of rigorous methods and standards used in base mapping.

A primary objective of data base preparation is to create a vertically integrated data base in which each feature is represented by only one set of coordinates. Features should be collected from the theme with the highest positional accuracy (boss layer).

Base cartographic layers will have a higher level of accuracy than resource layers. A hierarchy of positional validity should be established among the data themes to be acquired prior to data entry and adhered to during the digitizing process. For example, a stand boundary following a lakeshore must yield to and be coincident with the shoreline collected from the base layers.

Digitizing sequence**—**The cost in both time and resources dictates that what is to be digitized and when in the project sequence it should be done be considered early in all digitizing projects. Only essential data should be digitized, and they should be digitized in proper sequence. Units should refer to GIS needs assessments that identify resource layers required.

Digitizing sequence is important because correctly completing some layers depends on others of higher positional validity. Therefore, highest validity layers should be digitized first. PBS map layers have highest validity and should be available before resource digitizing commences. Validity assessment needs to be made for resource layers. Least positionally valid layers should be digitized last if some other layer controls polygon boundaries.

Map manuscript preparation**—**Before digitized data entry begins, it is often necessary to prepare a map manuscript as a guide for digitizing. Preparations for map manuscript production need to be carefully organized and monitored as the process develops. Map manuscripts need to meet rigorous quality standards before they are digitized or scanned and put in an electronic data base. Dull and others (1989) discuss the use of pin-registered Mylars, tic placement, pen colors, and feature labeling. Pen width and line connecting are very important, particularly if the data are to be scanned. A line on a map may actually be wider than the feature it represents. A pen size smaller or equal to the width of features or the distance tolerances for a particular scale should be selected, if possible. Jeweled or tungsten-tipped Rapidograph pens should be used, because Mylar is highly abrasive and destroys steel-tipped pens quickly. Lines of even width can be drawn with Rapidograph pens. Maps drawn with felt-tip pens are undesirable. Lines must touch where they are supposed to meet. Overshoots are not allowed, except when matching a positionally superior line.

Format refers to area coverage on a given map sheet. Most resource data will be mapped on the same format as PBS maps, i.e., 7.5 minutes of latitude and longitude (except for Alaska, where 15 minutes of latitude and 20 or 22 minutes of longitude are used). Some maps of data may not conform to this format. Nevertheless, all data should be digitized in (or reduced to) PBS format for ease of indexing and retrieving and storage consistency with other data.

Content editing is one activity that should be performed without exception, yet is often overlooked. Someone knowledgeable in the resource should review the layers to be digitized to assure completeness, correctness, and proper polygon identifiers.

Each PBS quad has eight neighboring quads. Lines of resource features do not end at quad boundaries; it is necessary to make sure that they continue onto adjoining sheets and that they match positionally with neighbors at quad boundaries. Matching and joining across edges are required for credible GIS's.

Each data layer that is subordinate to another of higher positional validity should be overlaid on its boss layer, and collinear polygon boundaries should be identified. Where such line segments are found, they should be copied "up" to the new theme layer from the superior positional accuracy "boss line" coverage. This simple step eliminates slivers and gaps that would otherwise be created. Experience has shown that 1 hour spent in this procedure will save about 4 hours of editing slivers and gaps from files.

Registration—A GIS consists of computer files of geographically referenced data. There are a number of candidate reference frames based on different mapping projections of the globe. The reference frame must be coherent and uniform over the extent of the area covered by the GIS. For a large-area GIS, the reference frame must be geodetically correct in order for data files from adjacent map units to match position and accurately represent actual ground positions. The reference system preferred by Federal agencies (the USGS as well as the Forest Service) for digital purposes is the UTM system, which is based on the North American Datum of 1927 (NAD-27). The UTM coordinates and distances are given in meters. The system is organized into north-south zones that are 6 degrees of longitude (east-west) wide. The zones begin at 180 degrees west longitude (the International Date Line), and are numbered in sequence, moving to the east. Fifteen UTM zones cover the area of Forest Service interest.

Another common reference system is the SPC. Although widely used for mapping and engineering, it has shortcomings for geographic data bases of large extent. The SPC measurement unit is the *American Survey Foot* (39.37 in/m). It differs from the more commonly used *International Foot* (25.4 mm/in) by two parts per million. Positional errors of several feet result when the wrong conversion factor is applied to large SPC coordinate values. Another vexing problem with the SPC involves the large number of zones (over 90 for Forest Service lands). An individual forest may lie in two or even three zones, each with separate coordinate system origins. Digitized data from adjacent zones must be transformed into a single system for forestwide GIS usage. This can cause confusion because of resulting alien values and will increase relative error.

At least four control points of known position in the chosen reference system must be digitized in order to establish the best fitting relationship of position, scale, and orthogonality (squareness) of the document with respect to the reference frame. A map to be digitized is properly registered (fitted) to the coordinate system when four or more points of known position used for control are digitized, and when correct coordinates for the proper system zone are entered into the computer for each control point. Computer programs to establish registration should employ a six-parameter coordinate transformation that will disclose how good the fit is, in terms of RMSE (see chapter 5). Typically, RMSE should be less than 3 meters ($\cong$10 feet) for 1:24,000 scale maps. Values significantly greater than this indicate a mistake, which should be found and corrected before proceeding further. The six-parameter transformation corrects for scale differences in original documents, allowing use of maps on either stable materials or paper. Rubber sheeting transformations should be avoided because they disguise blunders, excessively warp, and create artificially good fits.

Digitizing programs convert digitizer coordinates into ground units in the coordinate reference system. Because the program corrects for distortions in original maps, final ground coordinates of points are in correct relationship.

File structure and coding—Files created must have proper and explicit ties to attributes, through either tables or features links, and be topologically structured. Outputs should be examined to assure that these criteria are met.

Cardinal rule for digitizing:
Each line segment should be vector-digitized once and only once—both on and between layers. This rule must be followed to avoid redundant data that would have to be cleaned out of data files. This cleanup is time-consuming, tedious, and expensive.

Quality check—Digitizing quality is checked for content and position by obtaining plots of layers digitized and comparing with original documents.

Stable base film, not paper copies of original Mylar layers, is recommended as the preferred material from which to digitize. Because this material is dimensionally stable under changes in humidity and temperature, both content and positional checks can be made by comparing plots on stable base film with originals.

If the original document is paper, the plot of ground coordinates will not exactly overlay, because the six-parameter transformation corrects for paper shrinkage, stretching, and other distortions. Therefore, it is not possible to check for position quality after digitizing from paper maps with a plot on stable base film using ground coordinates.

Accuracy—Digitizer accuracy has become a subject of concern. The NMAS have been cited as standards toward which we must strive for digitizing. Unfortunately, there is much misunderstanding of the NMAS. Accuracy is important, but accuracy achievable by digitizing from maps (output) is *not* the same as NMAS (input). It is wrong to assume that output quality of digitizing can be equal to input quality of maps as constructed.

The NMAS call for 9 out of 10 *well-defined*, *checked* points to be within 40 feet (12 m) of their true position, at 1:24,000 scale (American Society of Civil Engineers 1978, Department of Defense 1981). An example of a well-defined point is a "T" road intersection or a bridge, *not* a winding stream or logging road. Misinterpreting this standard has led people to believe that all features shown on a USGS quad are accurate within 40 feet. A better guess might be twice that!

Some users assume that digitizer resolution (least count) is equivalent to digitizer accuracy, and that a digitizer displaying 0.001 inch (0.025 mm) can scale ground coordinates to 2 feet (0.61 m) from a 1:24,000 scale map. This assumption ignores inherent digitizer inaccuracies (e.g., perhaps the scale is not equal in x and y, or the axes are not perfectly straight or square). The real accuracy of many such digitizers is probably closer to 0.005 inch (0.127 mm), or 10 ground feet (3 m) at 1:24,000.

If one treats map position, fit to control, and digitizer errors as random errors, the "maximum probable" error of even well-defined, checked points digitized from a stable map with 0.005-inch digitizer error, registered to 2 meters RMSE, is 76 feet in x and y. The maximum probable error of natural features may easily be two or more times this amount. The digitizer component of this error is comparatively minor, indicating that concern over digitizer precision often is misplaced.

In view of this, excessive concern about proof-plot accuracy, as compared with originals, is unwarranted. Misfits of perhaps 0.030 inch (0.76 mm) may be tolerable, especially for resource themes that are not well defined in the first place.

Digitizing systems—Computer software systems used for digitizing should emphasize geodetic and cartographic concerns. They should permit easy transformation of coordinates from one system to another, solve registration functions rapidly, allow windowing, perform many edit functions (such as snapping, joining, and trimming) automatically, and process data from many digitizing sources. They should output data in a variety of formats to accommodate different GIS's. Digitizing maps showing resource information is a major task confronting agencies establishing a GIS, and the Forest Service is no exception. But it must be done before a GIS can be successfully implemented, and if it is not done thoughtfully, it will create future expense, trouble, and delays. Data entry is the major cost component of a GIS; we must be careful not to make an already expensive job even more costly.

Computer Spatial Data Bases

An electronic data base must have geographically referenced records if it is to be used in a GIS. Some data bases are spatial in that they contain information about the Earth's surface in relationship to other information, but there is no record that gives the location of the information on the Earth. Methods have been developed to reference some types of information, such as digital satellite imagery, but tabular data that were not referenced when collected may be difficult or impossible to reference.

Georeferenced data can be put into a GIS electronically, but they may have to be reprojected in order to match the location of data already in the GIS. Common map projections used in a GIS include UTM, SPC, Lambert Conical, Albers, and many others. Most GIS software packages have one or more projection functions that allow presentation of the data.

GIS information retrieval system types:
1. Multiple-attribute tables
2. First-generation relational data bases
3. Modern relational data bases using a structured query language

Three types of tabular systems have been used for spatial data in GIS's: multiple-attribute tables, first-generation relational data bases, and modern relational data bases using a structured query language. Data in multiple-attribute tables cannot be related to other data sets. Adding, updating, or deleting data may be difficult at best. Operands may or may not be available for using values in one item or column to calculate values for a new column. No operands were available in early versions of

the Mapping Overlay and Statistical System, for example, but were added to the software later. In relational data bases, data residing in different tables are related by having one column (a relate item) that contains record keys common to both data sets. Other columns contain distinct sets of information. Data from the columns in more than one table can be queried and displayed or used to calculate values for a new column; however, only a few mathematical operands are available, making complex modeling difficult. Data in data bases using a structured query language are also cross-referenced by a relate item or column. Query and display in these languages are more efficient, and a nearly unlimited number of mathematical operands are available. Complex operations using data in the related tables have become a reality in these advanced relational data bases.

Data structures for GIS's:
1. Vector data: line and polygon
2. Raster data: fixed cell
3. Quadtree data: variable cell

Three data structures are used in GIS's: vector, raster, and quadtree. Vector systems use a series of points, lines, and polygons with label points for the points and polygons. A raster system is composed of cells all of the same area and with attributes associated with each cell. Quadtree systems use cells of variable size. Of the three, raster systems require the most storage space, but are the easiest and most efficient to use in performing complex mathematical operations. Some GIS analyses can be performed with all three data structures, but the most complex analyses can usually only be performed with raster data structures.

High-quality commercial GIS's generally contain functions that will transfer information from public-domain GIS's and other public-domain data bases to a commercial GIS. Transfer between commercial spatial data bases may be possible between cooperating companies. Public-domain GIS's typically don't have functions for transfering data from one GIS to another, but may have functions for using other public-domain spatial data bases.

Inventory Data Bases and Reports

Up to this point, we have limited our discussion almost entirely to conversions of data to GIS data formats. How does this relate to corporate data base requirements for information? Corporate data bases may specify that mapped or inventoried information include a layer or number of layers of GIS information. They may require locational inventory information that is filed in report form. This information may be almost location-free, or it may have some fairly specific site information. Currently, the corporate data base is likely to be aggregated to levels that are not useful for inclusion in GIS's. Most managers have yet to recognize the data structures revolution in the microcomputer industry. Dispersed processing with a large central data base has not been the focus of recent developments. Communication between a corporate data base and a GIS will undoubtedly be possible within a short time, but trying to place all data management under a single central control, as in the past, is likely to impose new limits on thinking that are not productive in the long run.

Updating and Maintaining Existing Data

The currency of information is important to modern land and resource managers. Conflicts between owners and contractors can arise simply because adjoining inventories are not of the same age. Updating information and maintaining the currency of existing information have become increasingly important.

Aerial Photos and Imagery

The faster the landscape changes, the more quickly maps, digital data, aerial photography, and digital imagery become outdated. Scheduling the acquisition of new imagery involves tradeoffs between cost and utility of existing imagery. As the landscape changes, the utility of existing imagery for field navigation, inventory stratification, timber sale delineation, and many other applications decreases. At some time, the cost of acquiring new imagery is outweighed by the utility of new information for field activities and the improved accuracy of inventories.

Resource managers should make appropriate use of all available sources of imagery to meet user need. Medium-scale aerial photography is probably the single most useful imagery format for resource applications. Identifying the availability and recognizing the utility of alternative sources of imagery will insure that the most appropriate information is utilized to meet specific requirements. Alternative sources of imagery include national and regional aerial photography programs, satellite imagery, and site-specific aerial photography or video imagery. Satellite imagery, which costs significantly less on a per-acre basis than aerial photography, can be acquired on a biennial basis to complement medium-scale aerial photography, provided that human and computer resources are available to process this imagery.

The time period from flight request until products are available to users defines the minimum cycle for acquiring new imagery. For aerial photography, this includes the time to develop and process bids, to acquire the imagery, and to index, annotate, and print the photography. Where the photo acquisition time is limited to a relatively short period each year, 2 or more years may be required to acquire imagery of a management unit. In these cases, 4 to 5 years is probably the minimum interval between acquisitions of new aerial photography. For satellite sensors, imagery of an entire management unit may be acquired during a single 16- to 26-day revisit cycle, providing the area is cloud-free. Under these circumstances, it would be practical to schedule imagery acquisition on a semiannual or annual basis.

It is important that aerial photographs be properly maintained. Extremes of temperature and humidity and prolonged exposure to sunlight should be avoided in storing original negatives and transparencies as well as working prints. If prints will be ordered frequently, the agency or commercial processing laboratory is the best place to store the original imagery. Prints used in the field may be protected by using laminated or plastic holders. Imagery should be filed so that individual prints are readily accessible and returned after use. Damaged prints should be replaced, and users should be encouraged to order duplicates or enlargements for specific activities. Original copies of digital data should be stored separately to reduce the possibility of accidental erasure. It is important to develop procedures to ensure that derivative digital products can be associated with the original imagery.

All maps represent past conditions the day they are published. Thematic change is the primary reason for difference, but change in technological capability of describing and mapping can also be present. Before undertaking data base updating, one must determine the significance of change that has taken place since original mapping. What change is sufficient to warrant revision? Answers vary for each theme. PBS maps are reviewed on a cyclic basis, and change is incorporated where noted. Should thematic data be reviewed temporally or quantitatively? The answer depends on the nature of data and their relative importance. Soil maps, for example, should have a long "life expectancy" if competently produced in the first place.

Time frames suggested for cyclic review of maps:

- For *derived topographic layers* (such as slope and aspect), update whenever better elevation data are available.
- For *insect and disease surveys*, update monthly or yearly as the situation warrants.
- For *base maps,* revise cyclically. The established schedule is usually 5–7 years.
- For *vegetation* that has slow growth or infrequent modification, a 10-year cycle may be adequate. For vegetation that has vigorous growth or frequent modification, update yearly or up to every 5 years.
- For *soils and geology*, update every 25 years or whenever new information is available.

Many forested areas may have had some, if not all, of their area mapped into stands at some time in the past. In all probability, some data base records were tied to the stands as earlier mapped. A wholesale redelineation may make it impossible to use previously gathered field information or make it possible only with a lot of questionable "cobbling" of the associated data.

Timber harvesting, insects, disease, and fires can rapidly change the resource situation. Where these new conditions do not exist on imagery, the affected area may have to be traversed and the traverse plotted on the base map. The traverse should be done before beginning any new stand delineation, and then the stand boundaries should be transferred from the base map back to the photo. If many unmapped changes have accumulated on a large scale or in a sizable region, consider the use of satellite imagery like the SPOT-generated orthophotos. Note that imagery acquisition time will often make updating of this imagery necessary, too.

A first step in delineating new stands should be to transfer or extend the boundaries of stands in adjacent compartments into the compartment being mapped, where it is appropriate to do so. This provides a starting point that lends continuity to the stand map.

During use and especially during revision, it is inevitable that discrepancies in original data will be noted. These will generate questions of tolerable discrepancy before requiring corrective action. In a sense, these questions are similar to the question of actual change and probably should be addressed using the same significance criterion.

Most revision will be from imagery photogrammetrically transferred to map bases. For economy, the smallest photo scale practical should be used to expedite the photo transfer process. Doubling photo scale increases photogrammetric transfer work by at least four times.

Unbiased connection to original spatial reference is critical in the revision process. This connection is best performed using controlled (aerotriangulated) photography. Where available, the connection may typically be made to an accuracy of 10 feet (3 m) or better. Where controlled photos are unavailable, graphical connection is necessary. Location error will be greater, depending on scale and quality of graphical control used. For 1:24,000 scale, connection is typically 30 feet (9 m) or more. Positional discrepancies between original and revised data will never be better than this "reconnection" accuracy.

New positional location technology is rapidly emerging. Locational information referred to as the North American Datum of 1983 (NAD-83) is becoming increasingly available, especially data collected by GPS satellite receivers. NAD-83 is a better model of the Earth than the older NAD-27, which has distortions and inaccuracies, but to which all current mapping is referred. Coordinates can differ by several hundred feet, so conversion of data from one to the other system absolutely must be performed before GIS processing to maintain consistency of values. The public domain program NADCON performs this conversion to submeter accuracy and is available from the Department of Commerce, U.S. National Oceanic and Atmospheric Administration, National Geodetic Survey.

Computer Spatial Data Bases

Guidelines for protecting, updating, and maintaining spatial data bases must be carefully thought out and followed. Failure to do so can result in erroneous conclusions from data analyses and loss of data. Spatial data bases are easily corrupted, particularly when two or more persons are using them. This is especially true in a GIS when data sets are being manipulated, overlaid with other related data, merged with data from adjacent locations, or divided into data subsets for smaller geographic areas within that of the original data set.

Regular backups of data are essential. If data are not backed up, they will be lost sooner or later—guaranteed. No computer system has been designed that is free of system crashes. This is the most important concept of protecting data bases. Backups can be done periodically or sporadically. If data are being used and updated regularly, they need to be backed up daily. If they are being used daily, but few changes are being made, they can be backed up weekly or monthly. Data bases used infrequently can be backed up after each use or change.

Responsibilities for backups need to be assigned. Backups of data used by more than one person need to be assigned to a data base administrator. For systems where all data remain online all the time, the data base administrator can backup all the data at designated intervals. It generally works best if backups of data bases on removable media that are used infrequently are the responsibility of the primary user.

Data storage and work directories should be kept separate. Only the "owner" of a given resource data theme should have edit access to the directory where the original or official data are stored. For example, a wildlife biologist would own the wildlife theme, a range conservationist the range themes, a forester the timber themes, an archaeologist the themes for cultural sites, and so forth. The owner has the authority to update or correct records and certify that a given data set is correct. Users copy data to their work areas and use the copied data for GIS or other analyses. The official data would remain unchanged.

Because of the temporal nature of resource data, the owner would have to make periodic updates. The recommended procedure for this is to copy the old data set to a new one and archive the original as a historical record. The copy is updated and becomes the new, official record. If the data are such that a historical record is important, more than one copy should be archived.

A data dictionary describing each data layer or theme should be kept and updated each time a change is made to the data. Required items in the dictionary are (1) projection, coordinate systems and units, (2) electronic location, including archives and backups on removable media, (3) date of the last update, (4) scale at which the data were collected, and (5) a description of each file in the layer or theme, telling what it is. A narrative description can be used to give other information, such as the original date of the data, how and under what conditions the data were collected, accuracy and limitations of the data, problems with the data, and from where data may be obtained.

Data dictionary for data base management must include:

1. Projection, coordinate systems and units of measure
2. Electronic location
3. Date of the last update
4. Collection scale
5. Description of files by layer or theme

The data dictionary should be maintained by the owner of the data. Most GIS's do not now automatically track all required items. If some items (such as the reference projection) are logged in a header file for a particular GIS, these items may not need to be included in the dictionary.

Inventory Data Bases and Reports

Many recent inventories are now available in computer data base format. Some information that we are interested in may be protected for a variety of reasons. Especially sensitive for FIA units may be the location of plots. Cooperators in the East, where much of the land is private, are protective of their privacy and do not wish to have locations revealed. Nonetheless, information from these surveys may be localized to the county or occasionally subcounty level. Statistical techniques can be applied to apportion county-level data to mapped areas in rather sophisticated ways. For example, forest and nonforest land might be mapped and volume assigned to the area of forest in the county. If there were more detailed forest-type assignments in an existing GIS, then volumes might be assigned by forest type (Czaplewski 1990b).

Even printed reports of existing inventories might well allow summaries to be broken out by forest type within counties or other administrative boundaries, although the effort and expense of doing so could be prohibitive. Later in this chapter, we will discuss some methods for updating existing inventory results that could aid in the utilization of these older printed reports.

Spatial Information from Different Sources

Data often result from surveys or inventories whose designs were not optimized for the particular variable or data element. When compared to the information from an inventory with more specific information regarding that variable, such data may lead to conflicting information. For example, when the area of forest land in the United States was estimated by the Forest Service and by the SCS, the two reports disagreed, resulting in a dispute over who was right. In fact, the conflicting estimates were due to three factors: (1) definitions differed significantly (for example, the Forest Service classified land covered with pinyon-juniper stands as forest, and the SCS called it woodland); (2) numbers were not compared with their sampling error attached, because a meaningful statistical comparison of the two estimates was rejected in favor of an accounting type of comparison; and (3) the ages of the estimates were actually different for different areas of the country, at least partially accounting for local differences within regions of the country.

In the following sections, we will try to present a more reasoned approach to dealing with estimates of a single resource from different sources. We present this in somewhat more detail than some other subjects because it may not be generally available in the wider GIS literature.

Combining area statistics from independent samples—Suppose that the Enchanted Forest needs estimates of area and stand conditions for three types of forest land (hardwoods, conifers, and open brushland, differentiated by low, medium, and high densities and by stand age classes). These estimates are required for forest planning or more spatially detailed planning that considers cumulative impacts of multiple projects. Area estimates will be used in models that predict future distribution of stand conditions for timber and wildlife assessments. Area estimates also are used frequently for general statistics to describe the current status of the forest. At the outset, let us recognize that no two surveys will obtain exactly the same result for estimates. Recognizing that two independent estimates of the same parameter are stronger than a single or even repeated estimate using the same method, we wish to obtain the best estimates we can.

The Enchanted Forest is indeed blessed with two different sets of estimated percentages for various types of forest cover. (We also assume that the two data sets, unfortunately, were not collected as part of a two-stage sample, with the advantage of statistical estimation that this would have provided.) Data for each of the two sets were gathered at about the same time. The first set comes from remotely sensed thematic maps (e.g., digital classification of Landsat) and the second from a randomized sample of 1-acre (0.04-ha) field plots (e.g., timber inventory). The sum of estimated percentages equals 100 for each set of the two sources. However, the estimated proportions for specific cover types differ. For example, the remotely sensed maps estimate that 56 percent of the Enchanted Forest is forest, while the

sample of field plots estimates that only 51 percent is forest. The inconsistencies between the two sets of estimates are even larger when categories of forest cover are compared. People usually tend to wonder which set of estimates is correct, or whether either is correct or better than the other. In fact, such questions are probably counterproductive. A better question to ask is, "How do we combine the results from these two independent estimates to obtain a better estimate?"

We have already discussed misclassification bias in the remotely sensed estimate of forest (see chapter 5). This will have to be considered while combining information. The statistical estimates from the sample of field plots are unbiased (i.e., by definition they do not contain misclassification bias), but the field estimates of area in various forest types have associated sampling errors. Field examinations are expensive, and only a relatively small sample size of field plots can be measured. The difference between the field estimates and the true, but known, forest areas is part of the random selection process for field plots. (Remember, you would not be surprised to get six tails when flipping a coin 10 times instead of the expected five tails).

Neither the remotely sensed nor field estimates are exactly correct, but both estimates are probably "close" in some sense to the true percentages of forest in various condition categories. As a first step, you might want to test to see if the two answers are the same in the statistical sense. There might be different levels of classification detail in remotely sensed estimates from thematic maps and sample estimates from field plots. For example, the percentages of a forest type in different age classes might be estimated with field data, but stand age might not be satisfactorily classified with multispectral satellite data. How can the field data, with high thematic detail but low spatial detail, be used with the large amount of data from remote sensing, which has high spatial detail but less thematic detail, to produce area estimates?

Statistically combining the remotely sensed and field sample estimates into a better estimate provides the resolution to the two-answer conundrum. As a first guess, you might take the average of the two estimates for percent forest: (56 percent + 51 percent)/2 = (56 percent × 0.5) + (51 percent × 0.5) = 53.5 percent. The second form of notation shows that averaging can be thought of as weighting. Weighting here implies that you have the same confidence in each estimate and, therefore, put the same weight on each estimate (i.e., 0.5).

But in our example, the number of plots (sample size of field plots), is small (i.e., high sampling error), and you expect that misclassification bias is relatively small. In this case, you might expect the remotely sensed estimate to be closer to the true (but known) percent forest than the field estimate, and you might put more weight (e.g., 0.6) on the remotely sensed estimate and less weight on the field estimate (e.g., 0.4); the combined estimate of percent forest would be (56 percent × 0.6) + (51 percent × 0.4) = 54.0 percent. (The difference between weighted and unweighted is small in this case. This might be considered "splitting hairs," but the choice of weights would have greater effect with more detailed classification than you have for forest type.) The selection of 0.6, however, was arbitrary. The question remains, what objective criterion should you use to determine a defensible choice of weights?

The statistical composite estimator provides the answer (Schaible 1978). Each of the two independent statistical estimates is weighted inversely proportional to its variance. The more precise estimate (i.e., the estimate with less variance or a smaller confidence interval) would receive more weight than the less precise estimate. Gregoire and Walters (1988) note that composite estimators are widely used in forestry, including sampling with partial replacement (e.g., Ware and Cunia 1962). Green and Strawderman (1986) and Thomas and Rennie (1987) show how a composite estimator may be used to combine independent estimates of stem density, basal area, or wood volume.

The first objective of this section is to give a simple example that shows how a composite estimator is applied. For the sake of simplicity, the example focuses on estimation of percent forest. However, more detailed classifications of land cover will usually be required in practice, using multivariate composite estimation. The example below can give you an intuitive feel for how composite estimation works. However, multivariate composite estimation should be applied by someone with statistical experience with this technique; otherwise, biased estimates might be unknowingly produced.

Sampling and misclassification errors cause our estimates to differ from the true state; these errors can be minimized but not eliminated. Unfortunately, another type of error is often encountered in practice: forest and land cover change over time, and these changes cause biased errors when old data are used to estimate current conditions. If the magnitude of change can be predicted to "update" the older estimate to current conditions, then past data have value, even in the presence of change. The second objective of this section is to give a simple example of combining old statistical estimates, model predictions of changes in forest cover, and low-precision current measurements of forest cover (e.g., remote sensing or field plots).

Making a composite estimate—First, consider the problem of estimating the percent of forest cover on the Enchanted Forest.

A total of nine photo interpreters independently made ocular estimates of forest cover on the Enchanted Forest. Results are recorded in table 9. The mean percentage (x_1 = 56.0 percent, table 9) is the first estimate of percent forest cover (see figure 10), with variance of the mean $Var(x_1)$ = 3.11 percent percent (variance units are percent^2, denoted as percent percent). Using simple statistical formulas and 8 degrees of freedom (df), this produces the approximate 95-percent confidence interval:

$$\mathrm{CI}_{.95}\,(x_1) = 56.0 \pm 3.5 \text{ percent.}$$

The variance in this example is intended to represent the variance in a calibrated, remotely sensed estimate.

Table 9—Independent estimates of forest cover, Enchanted Forest

Observer	*Ocular estimate x (percent)*	x^2
1	60	3,600
2	55	3,025
3	50	2,500
4	55	3,025
5	52	2,704
6	50	2,500
7	65	4,225
8	62	3,844
9	55	3,025
Total	504	28,448
Mean	56	–

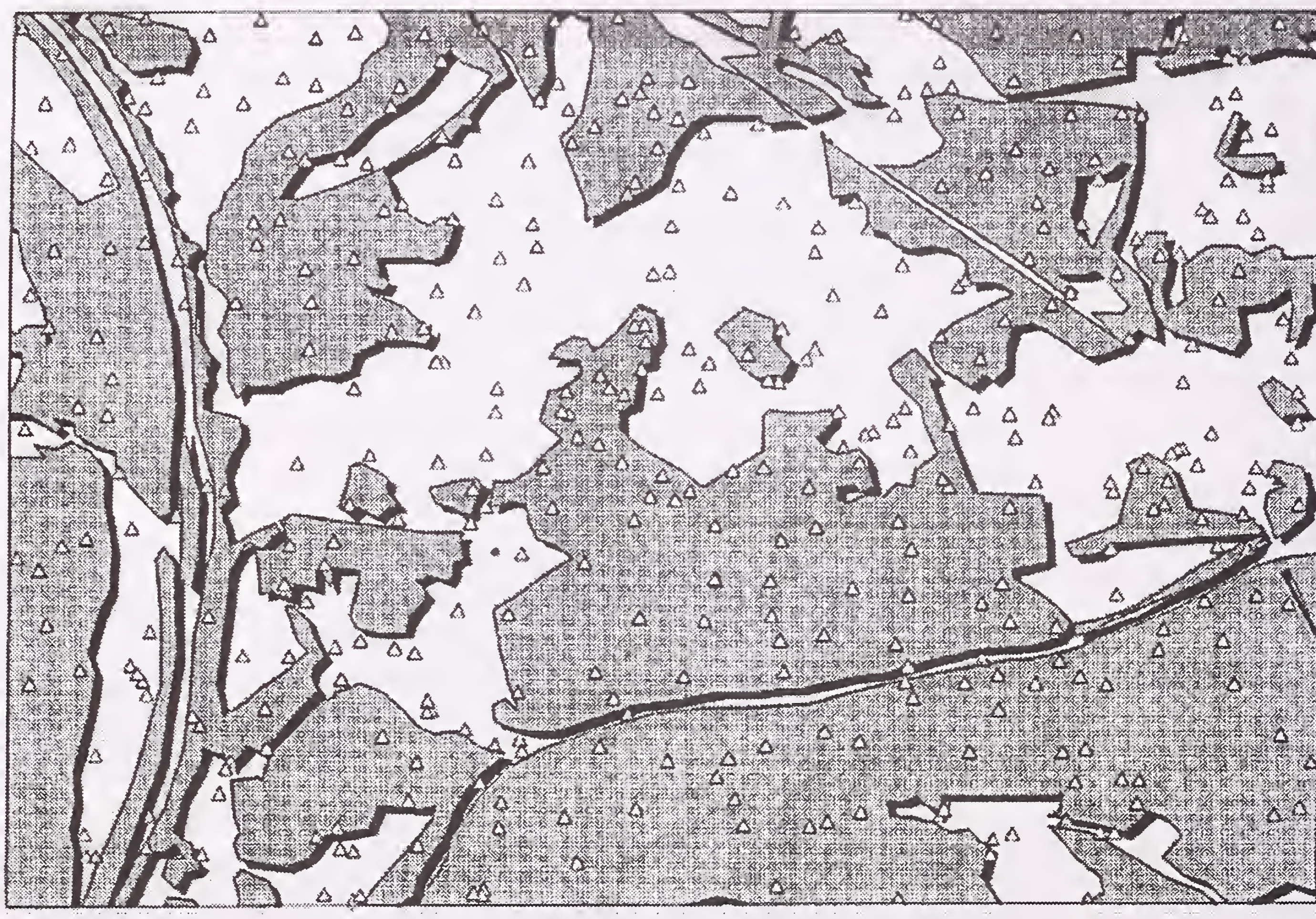


Figure 10—Estimate of the percent forest cover (shaded area) on the Enchanted Forest at time t = 1 (triangles are plot locations).

Next, consider a second estimate of percent forest cover shown in figure 10 using "error-free" classification of 400 temporary ground plots. There are 204 forested plots, producing the estimate y_1 = 100 percent(204/400) = 51.0 percent. Because we are actually dealing with proportions, we can use estimate of variance for the binomial distribution, $Var(p) = p\ (1 - p\ )$. The estimated sampling variance for the proportion $Var(y_1 = Var(p)\ /\ n)$ is:

$$Var(y_1) = (204/400)(196/400)\ /\ 400 = 6.25 \text{ percent percent}$$

from which, once again using simple statistical formulas, we can calculate the approximate 95-percent confidence interval:

$$\mathrm{CI}_{.95}\ (y_1) = 51.0 \pm 4.9 \text{ percent}$$

The estimate x_1 = 56.0 percent from table 10 can be combined with the field sample estimate (y_1 = 51.0 percent) to produce a new composite estimate x^*_1; as shown in figure 11, x_1 is weighted more heavily than y_1 because $\mathrm{Var}(x_1)$ = 3.11 percent percent for replicated ocular measurement error is less than $\mathrm{Var}(y_1)$ = 6.25 percent percent for sampling error from 400 point plots:

$$x^*_1 = [A_1\ y_1] + [(1 - A_1)\ x_1]$$

and we compute:

$$A_1 = \mathrm{Var}(\mathrm{x}_1)/[\mathrm{Var}(\mathrm{x}_1)+\mathrm{Var}(\mathrm{y}_1)]$$

$$= 3.11 \text{ percent percent} / (3.11 \text{ percent percent} + 6.25 \text{ percent percent}) = 0.33$$

and:

$$\begin{aligned} x^*_1 &= [(0.33)\ 51.0 \text{ percent}] + [(0.67)\ 56.0 \text{ percent}] \\ &= 54.4 \text{ percent} \end{aligned}$$

The use of the weight A_1 above is objective. It can be supported using a statistical optimality criterion (i.e., maximum likelihood or minimum variance estimation). The expected variance of the composite estimate $Var(x^*_1)$ is:

$$\begin{aligned} Y &= (a\ X_a) + (b\ X_b) \\ Var(Y) &= a^2\ Var(X_a) + b^2\ Var(X_b) \end{aligned}$$

Applying this theorem to the composite estimator x^*_1:

$$\begin{aligned} Var(x^*_1) &= [A_1^{\,2}\ Var(y_1)] + [(1 - A_1)^2\ Var(x_1)] \\ &= [(0.1089)\ 6.25 \text{ percent percent}] + [(0.4489)\ 3.11 \text{ percent percent}] \\ &= 2.08 \text{ percent percent} \end{aligned}$$

This estimator of $Var(x^*_1)$ can be found in any statistical discussion of weighted estimates, such as Gregoire and Walters (1988). The variance of the composite estimate is smaller than either of the two independent estimates, as illustrated in

figure 11; the approximate 95-percent bounds on estimation error for the composite estimate x^*_1 are $\pm$2.8 percent, compared to $\pm$3.5 percent for the mean ocular estimates, and $\pm$4.9 percent for the sample estimate using 400 plots.

Figure 11 shows expected probability densities for estimates of percent forest cover on the Enchanted Forest from mean ocular estimates (x_1 = 56.0 percent) and 400 point plots (y_1 = 51.0 percent). These are weighted inversely proportional to their variances, and combined into the composite estimate (x^*_1 = 54.4 percent).

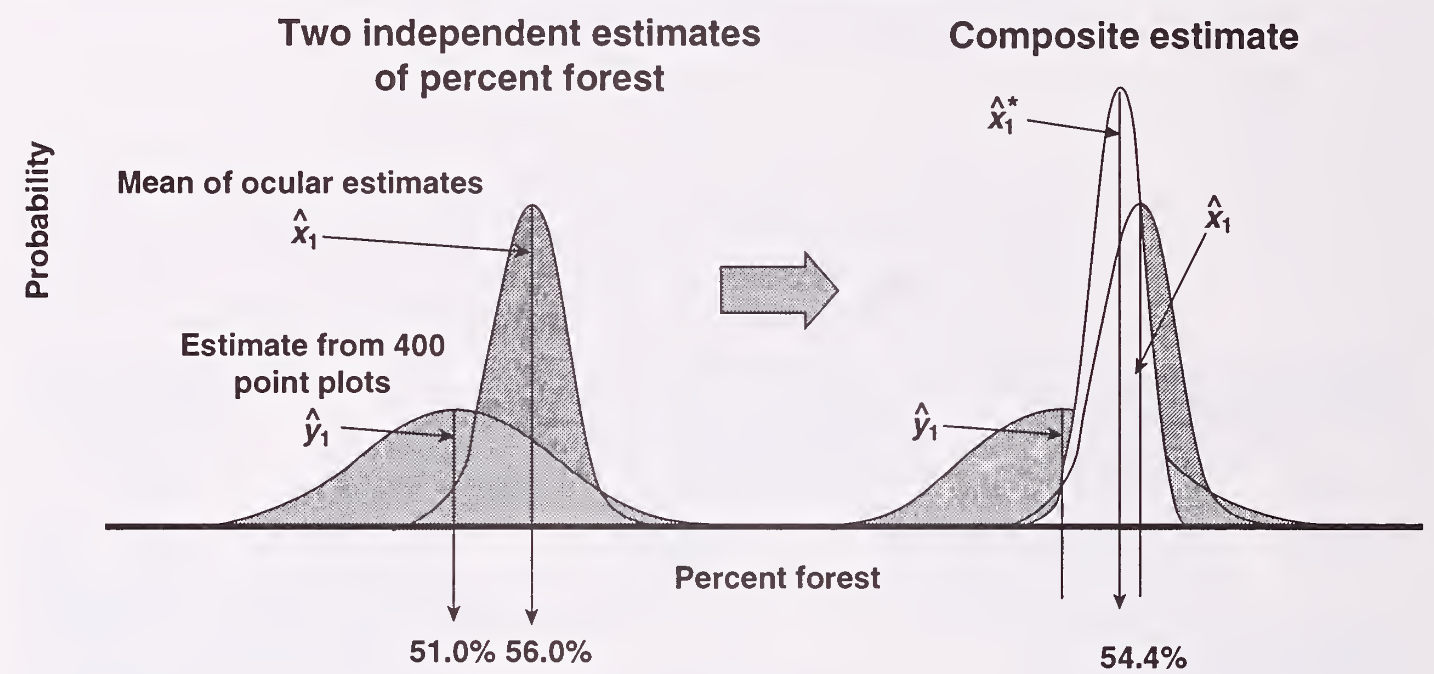


Figure 11—Expected probability densities for percent forest cover on the Enchanted Forest.

Estimating changes over time—To illustrate application of combined estimators to the Enchanted Forest, we present figure 12, which shows the condition of forest cover at time $t = 2$, after some disturbance has changed the conditions shown in figure 10 at time $t = 1$. First, make another ocular estimate of percent forest cover in figure 12. Then, record your answer next to figure 12.

In figure 12, there are 200 temporary plots independently classified to estimate percent forest cover ($t = 2$). Mean percent forest cover = y_2 = 45.0 percent, $Var(y_2)$ = (45.0)(55.0)/200 = 12.38 percent percent, and $CI_{.95}(y_2)$ = 45.0 $\pm$ 6.9 percent.

An estimate of percent forest x_2 can be made from the prior ($t = 1$) estimate x^*_1 = 54.4 percent, together with the estimated rate of change. Loss of forest class between $t = 1$ and $t = 2$ is estimated at 5 percent of all forest cover in figure 10. Therefore, the percent of stocked forest at time $t = 2$ is $x_2 = 0.95(x^*_1)$ = 0.95(54.4 percent) = 51.7 percent.

Here, we use the simple estimate of variance. For the linear transformation x_2 of x^*_1, the variance of the error at $t = 2$ is simply:

$$Var(x_2) = (0.95)^2 Var(x^*_1) = (0.90)2.08 \text{ percent percent} = 1.88$$

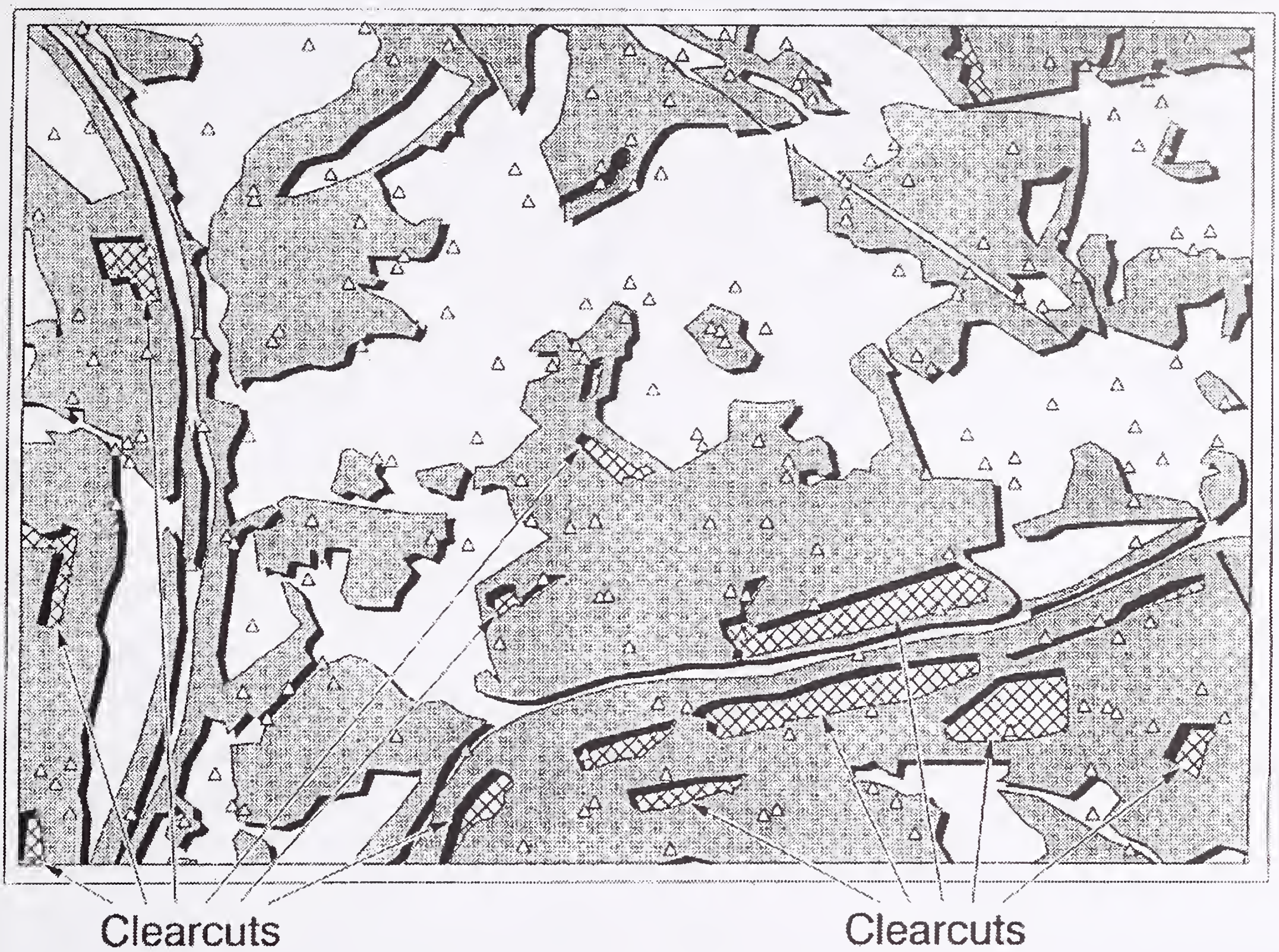


Figure 12—Estimating forest cover change between inventories (triangles are plot locations).

However, in situations where value is high in terms of dollars, time, or resources, we should recognize that models are imperfect. Prediction errors, denoted, occur; and a statistician should be engaged to apply the more sophisticated stochastic, linear transformation model $x_2 = 0.95x^*_1 + w$. In many cases, the expected value for w will be zero, so the updated estimate x_2 is unaffected, though the variance estimate will reflect the additional error term.

If prediction errors for w are independent of estimation errors for x^*_1, then $Var(x_2) = (0.95)^2\ Var(x^*_1) + Var(w)$. If $Var(w)$ is assumed to be 1.00 percent percent, then $Var(x_2)$ = 1.88 percent percent + 1.00 percent percent = 2.88 percent percent, which yields an updated estimate of x_2 with $CI_{.95}(x_2) = 51.7 \pm 3.3$ percent.

Figure 13 shows the estimate of probability densities for percent forest cover at time $t = 2$, based on information from $t = 1$. The model $x_2 = (0.95)x^*_1 = 51.7$ percent (i.e., 5 percent of the forest has changed between $t = 1$ and $t = 2$). Given this information, only the estimation error at $t = 1$ is propagated to $t = 2$. An independent estimate y_2 is available from the 200 plots in figure 12.

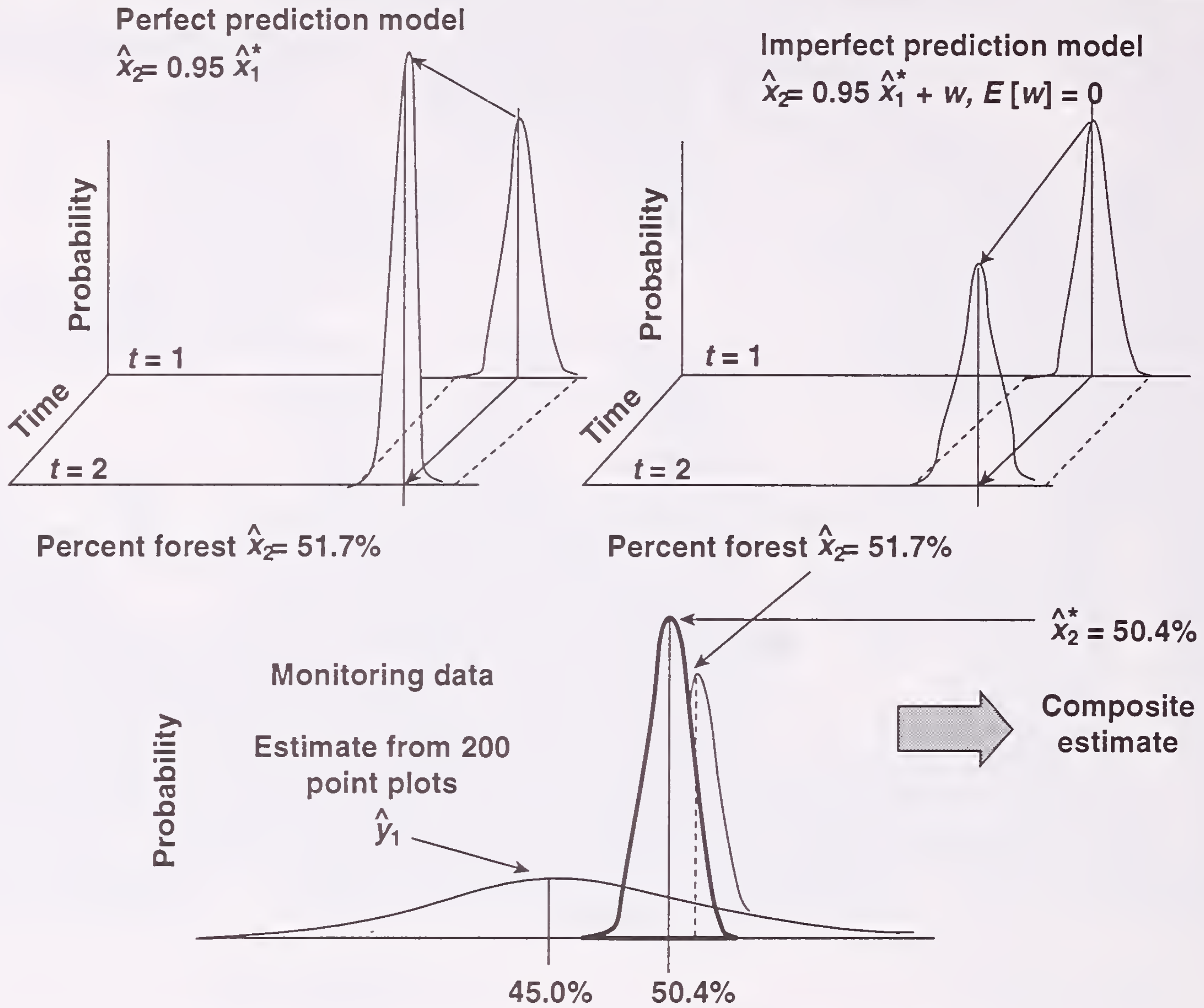


Figure 13—Estimate of probability densities for percent forest cover at time t = 2, based on information from t = 1.

Monitoring data over time—Monitoring resources through time is becoming increasingly necessary. Monitoring is the natural extension of combining two sets of data from sequential time periods. In the increasingly complex time series sampling situations we face, we are blessed with increasingly flexible statistical tools to model and analyze these time series. One common statistical method is recursive least squares (a basis for Kalman filtering[a]), which allows the combination of data from more than three sampling periods (Young 1984, Czaplewski 1990a), although it improves in value when there are more than 12 time periods. With each new monitoring measurement, a "composite" estimate is made, which serves as new initial conditions for the next deterministic prediction (see, for example, year 4 in figure 14). Monitoring measurements can adjust predictions from a simple linear model for trends that are truly nonlinear, but are not well quantified. Precise data from the past can improve current estimates using less precise, but more recent, monitoring data. The actual application of the Kalman filter can combine monitoring data from many sources, such as temperature, rainfall, and tree-ring series, but requires rather precise knowledge of variance-covariance properties of the system of equations. Estimation using recursive least squares has become increasingly popular; however, you should discuss its application with a statistician or two!

[a] Unfortunately, the use of recursive least squares has been referred to as Kalman filtering. It is related, but the Kalman filter can combine multiple time series and usually requires extensive information in terms of a state space and a transition matrix, which we do not need in order to update simple information vectors.

Figure 14 shows an example of recursive least squares estimates and approximate 95-percent confidence intervals. Forest inventories were conducted in years 0 and 10; monitoring data were gathered in years 4 and 7. A time series of relatively imprecise (i.e., inexpensive) monitoring data can prolong utility of a previous, more expensive forest inventory.

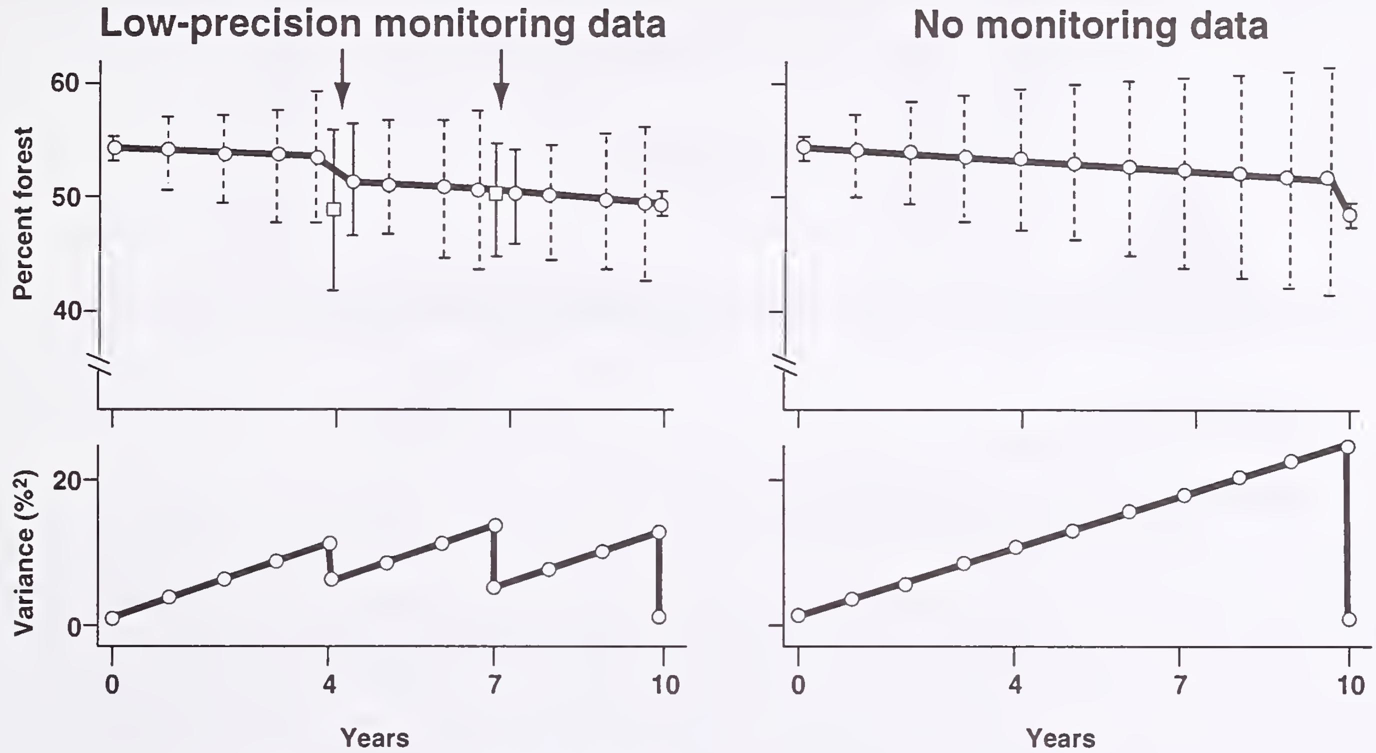


Figure 14—Recursive least squares estimates and confidence intervals.

Verifying divergent estimates—It is possible that two independent estimates disagree, or "diverge," in that neither estimate is likely, given the other; that is, their respective confidence intervals have a low probability of containing the other estimate (figure 15). Divergent estimates are cause for reexamining the sources of data. The possibility of blunder or nonstatistical error is more likely. However, real change over long remeasurement intervals may also result in nonoverlapping confidence intervals. The types of statistical problems that might account for divergent estimates are related to estimating the error distribution of (1) the current monitoring measurement, or (2) the past estimate that is updated by a deterministic prediction model with insufficient effort to estimate the variances. Nonstatistical problems sometimes can be corrected by reexamining the data. Statistical problems may be solved, but it is necessary to involve highly qualified statistical help. Remember, when estimates diverge, they still can be combined, but the results may be biased for the context for which we wish to apply them.

Figure 15 shows the predicted probability densities for two independent estimates (measurement y_t and model prediction x_t) that disagree; the residual difference between the two estimates makes it unlikely that they are equal. Combination of these two estimators is not recommended.

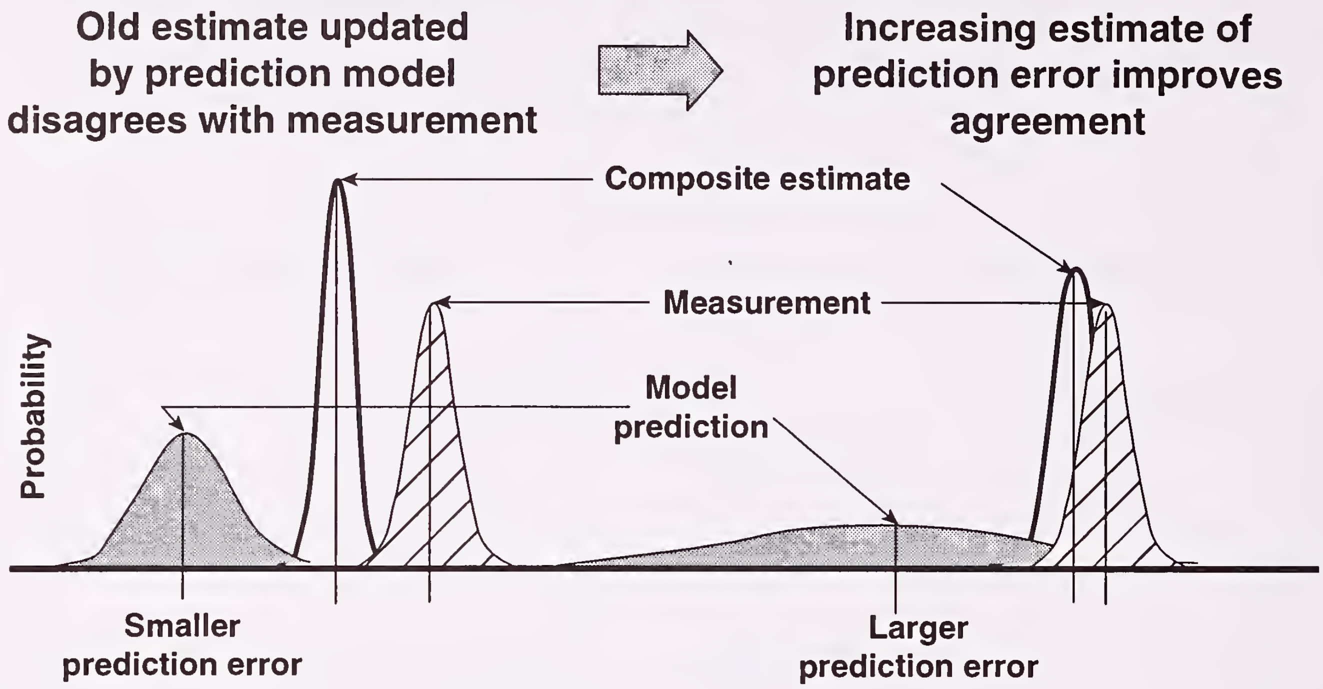


Figure 15—Predicted probability densities for two independent estimates.

Combining data from mutually exclusive areas (aggregation)—Aggregated data should be treated similarly to a single area with two estimates. This is not readily apparent, because with aggregated data we do not have conflicting estimates to compare. Nonetheless, a little reflection will reveal the similarity: just as two adjacent photos or maps, for example, may have different data time values, so different inventory areas may have different time values. Problems result from simple time differences when, for example, a 15-year-old inventory is to be combined with a newly inventoried area. Particularly egregious problems may appear when artificial strata occur; salability limits can pose serious problems. Suppose that you are interested in estimating growth of your salable timber, and that a 5.0-inch (12.7 cm) dbh lower limit exists. Cutting has been extensive, but not regular, so that for reasons of social economy many stands were cut since the last inventory, but most were cut immediately after. Now, when your estimate of growth is made, many regenerated stands are barely below salability. It is entirely possible that an estimate of growth would be seriously below the potential for the inventoried area. Now, combine this estimate with growth estimates from contiguous areas that did not have heavy cutting, and the overall estimate may be seriously biased. Both of these problems suggest that we should consider modeling the time difference and using estimators that consider the variance of the estimates before aggregating them.

Combining data from overlapping areas—Earlier discussion was meant to cover an area in which the boundaries were unchanging over time and entirely coincident in space. Often, inventories result from management regimes or programs that cover part of an area, but some may be entirely new, whereas other portions of an area are left out. The main consideration is that part of the area may have different variance of estimates. Breaking out (disaggregating) part of the new inventory and applying it in some organized manner will result in more useful information.

Disaggregating data—Disaggregation means breaking larger sampling areas into smaller ones. When this is done, the sampling error for the smaller area is considerably larger than for the larger one. If the original plot data are available, an estimate of the variance can be computed for the newly formed regions. Often, at least in the initial stages of establishing a GIS, we may have estimates that do not have original plot data. If there is at least an existing estimate of the standard error or variance, it is possible to get a first approximation of the variance from simple formulas and plot number assumptions. A normal approximation to the sampling error for the smaller area can be written as:

$$SE_d = \frac{SE_t\sqrt{X_t}}{\sqrt{X_d}} \qquad \textbf{(k)}$$

where the subscripts t and d represent total and disagreggated subtotals for the area and its subarea, respectively.

Computation of a sampling error allows us to combine estimates for inventory elements that may not be completely distinct or those that are overlapping. Once again, there is good reason to involve a statistician in the process.

Extrapolating data—Extrapolation, as used by natural resource managers, refers to estimating means or totals for areas in which no field data collection was made. In the statistical sense, extrapolation means estimation beyond the range of data. For general GIS application, "supplying missing values" may be a more appropriate description of this procedure than extrapolation. Applying data from one area to another is risky and unlikely to satisfy usual requirements for the forest GIS. However, there may be no alternative in some instances, and we should use procedures that have some valid basis in research where possible. Among the questions we might ask are: What kinds of similarity measures allow for extrapolation (supplying a missing value) to stands for which there are no field data? How far in terms of distance can we be from the derived stand data? Can we use old photos with relevant characteristics to estimate currently observed stand conditions (extrapolation through time)? Statistical procedures exist that can be used to help extrapolate data, among them regression estimators and missing value procedures (Kish 1965).

Using data with unknown sampling error—There are cases where a sampling error is simply not available. Where possible, the principle use of data with unknown sampling error should be for the purpose of allocating some sampling to establish an error limit. But it may be possible to find similar problems for which sampling error may reasonably be extended to the current data. Remember that sampling for an estimate of variability requires significantly larger samples than for the estimation of a mean.

New Mapping and Inventory Projects

The efficiency of any inventory or land mapping effort increases rapidly with the inclusion of existing information. Sample allocation schemes can be improved, and precision and accuracy can be increased. One strategy for using existing information is to allocate samples more heavily where change is suspected to be most. Even

in the determination of landform, using existing maps and photos might be helpful in determining which areas within the new mapping effort area need additional coverage. For example, an important watershed might have been cloud-covered in the photography available to the previous mapping effort. Reviewing the dates and character of the existing photography used to map landform is effective in allocating a new photo mission to obtain cloud-free coverage, and the rest of the existing photography can be used as is. Existing aerial photography and digital imagery can be used effectively in definition, development, and implementation of mapping and inventory projects.

Aerial Photos and Imagery

Aerial photography has been the principal source of information on the Earth and its features since the availability of the aerial camera platform in the late 1800's. Most standard series planimetric and topographic maps, DEM's, and cartographic data sets are based on aerial photography. Much the same holds true for delineations of resource data themes. Since satellite imagery first became available in 1972, it has become an increasingly important tool in the creation of both base maps and especially resource maps and inventories.

Natural resource inventories, mapping, and modeling are performed in response to the issues and concerns identified by resource managers and the public. The use of existing imagery can clarify and focus the decisionmaking process that defines requirements. Each individual involved in identifying requirements brings preconceived perceptions of the situation on the ground to the discussions. The availability of appropriate imagery can focus discussions and provide a reality check for participants.

Imagery can be an effective tool in the design of inventories and mapping activities. In mapping projects, current aerial photography can be used along with field visits to develop criteria, standards, and procedures. It can help answer a broad array of questions, including what vegetation types are present, what the minimum mapping unit should be, and how difficult it will be to traverse the area. Imagery can be an important tool for inventory design. In the simplest case, we can use imagery to subdivide the area of interest into strata, each of which is internally more homogeneous than the area of interest as a whole. Field measurements are almost always the most expensive part of an inventory. Imagery can be used to reduce the requirements for field sampling for a given level of precision or to increase the level of precision for given levels of sampling.

Multistage sampling has been discussed by numerous authors both within forestry and in other fields. References, discussion, and examples are presented in Lund and Thomas (1989). Key references relating to the use of satellite imagery include Langley (1975) and Poso and others (1987).

In mapping individual forest stands, delineation may be done either on aerial photographs prior to field examination or in the field. Aerial photography has been used extensively in the compilation of soil surveys for locating field sample points and as an important aid in delineating mapping units.

Aerial photography at a wide range of scales is used to locate and characterize wetland habitats in the National Wetlands Mapping Program. In this program, field examination serves as a check on photo interpretation rather than as the primary procedure for delineating mapping units. Aerial photography is an essential element in the design of resource inventories conducted nationwide by the Forest Service. Satellite navigation is regularly used as a base map and a basis for delineating resources in resource surveys of developing countries. Satellite imagery is being used to develop vegetation distribution maps on a national scale.

Aerial photography and satellite imagery can play an important role in all phases of resource inventory and mapping projects. Inventory activities are planned based on our perceptions of information needs and the area of interest. Our assumptions regarding the area of interest shape decisions in the design of inventories. An examination of available imagery may confirm the extent and condition of the resources and accessibility to them. Such information feeds directly into inventory design and execution.

Maps and Overlays

Of many recent technological advances, four that occurred within roughly the past decade provide significant opportunities for land management. Happily, a synergistic relationship exists among them. The four are wide coverage of high-altitude photography, deployment of GPS satellites, maturing of analytical photogrammetry, and development of a GIS. The first three of these technologies provide an ideal foundation for the fourth (Valentine 1990).

The use of GPS receivers is a superb means of providing geodetic control for large blocks of small-scale high-altitude photography. This controlled photography can be processed by analytical photogrammetry to provide a splendid source of quality, low-cost data suitable for a project GIS. The process yields highly accurate, three-dimensional positions of *every image* on such aerial photos! Therefore, every photo becomes a source of accurate geodetic coordinates of a virtually unlimited number of points and objects. Positional accuracy is ample for nearly all practical land management needs, reducing or eliminating requirements for ground measurement to gather project-level data.

Experience with GPS control and analytical photogrammetry demonstrates results better than anything possible with traditional methods. The Forest Service's R-6 has extensive experience with this technique, having controlled more than a dozen such blocks of 1:40,000-scale photography by GPS. The average fit of the analytical bridges is less than 2 feet (1 sigma) in NAD-83. This is abundant accuracy for a project GIS, well suited for all but a few very special cases.

A few thousand dollars will pay for controlling a block of high-altitude photography using GPS and analytical photogrammetry to cover an entire national forest. Proper use of analytical photogrammetry bundle adjustments and intelligent control positioning is very cost-effective. For example, horizontal control for a square 200-photo block is possible with just 20 judiciously placed stations. This means that we can provide control for 100 PBS quads (over 5,000 square miles or 12,900 square kilometers) covered with 1:80,000 photography (or 25 quads with 1:40,000

photos) for about $20,000. Forests can easily save in excess of $70,000 over conventional methods of control and placement.

In 1983, Potlatch Corporation of Lewiston, Idaho, digitized 81 quads covering its Idaho timberlands at a cost of only a few cents per acre. Using 1:80,000 photography and analytical photogrammetry, they developed a highly accurate GIS data base of roads, streams, ridge lines, property lines, and other important features. Potlatch foresters now use this data for local project planning of harvest and other operations.

Forests now have high-altitude photo coverage at 1:80,000 scale, and many are getting coverage at 1:40,000 scale under the National Aerial Photography Program (NAPP). The only missing link is GPS control. A typical forest covered by 50 to 70 quads could easily and economically get necessary control using GPS receivers. Personnel trained in cadastral techniques can help with this phase.

Several Forest Service regions have capability and capacity to perform analytical photogrammetric digitizing for those who want to make investments in quality data for a project GIS. Forests should acquire these data on selected areas of high value and activity requiring intensive planning. Regional geometronics leaders can make arrangements to have this work accomplished.

This aerotriangulated photo block is a ready source of control points (images) for other uses, virtually ending geodetic control needs (even the need for additional GPS readings). With this source, users can control larger scale or newer photography for projects needing ground measurements or surveys. With analytical plotters, one can measure to any practical level of *relative* accuracy by merely using photos of different scale. For example, about 1-foot (0.3-m) relative accuracy is possible from 1:24,000 photos, and a relative accuracy of about 5 inches (12.7 cm) from 1:10,000. Another advantage is that measurement accuracy is homogeneous. All features visible on the photo are measurable for whatever purpose, including a project GIS.

Future projects will be performed better from lessons learned through use of current data. We will gain insights on what is really critical, what errors are tolerable and what are not, and what data are actually needed and to what level of accuracy for specific purposes. For example, if you are going to update maps using existing information, first you must know the reliability of the existing information. Then you must evaluate the mapping procedures for this information. After taking these steps, you can start updating the information based on the results of the reliability and evaluation procedures used.

Computer Spatial Data Bases

Existing information in computer spatial data bases can be used to improve new mapping and inventory projects, particularly with the overlaying and other analytic capabilities in a GIS. There are numerous applications, with possibilities limited only by the functions in a GIS and by the user's knowledge of them. Several examples are given below to show how existing information can be used in new mapping and inventory projects.

Planning timber harvests—Reflected energy values from satellite imagery are reclassified to values representing several timber types. The types are attributed as to height and canopy closure, two items used for classifying areas for elk or deer thermal cover. Some ground truthing is done to verify the results. A third item is area, which is calculated in the GIS. The three items are overlaid in the GIS along with elevation data. Maps showing elk and deer thermal cover are created. The inventory of elk and deer thermal cover can be completed in a relatively short time because manual classification is avoided. The maps of thermal cover are used to show where timber harvest and thinning activities can occur without adversely affecting big game wintering areas. The cover maps can also be overlaid with foraging areas. The resulting maps are used to show where timber harvests can be planned to improve cover-to-forage ratios.

Protecting stream quality— Protecting stream quality is an example of buffer analysis. The existing information needed is stream classification (first-, second-, and third-order), timber volume data, and contour lines or DEM's from which to calculate slope. First-order streams are protected with a 100-foot (30-m) buffer, second-order streams with a 60-foot (18-m) buffer, and third-order streams with a 30-foot (9-m) buffer. The buffers are horizontal and therefore affected by slope. The steeper the slope, the wider the buffer zone uphill from the stream. The timber in the resulting buffer zones can be removed from the inventory for the annual allowable cut. All of the analysis is done in the GIS; new timber volumes are calculated and maps of the buffer zones drawn in an office setting in a matter of hours as opposed to months for field analysis and adjustment of timber volumes. With tree height information, we can take the analysis a step further. Tree height affects the amount of shading on the stream and thus affects stream temperature. If stream temperatures are marginally too warm for rearing anadromous smolts, selective cutting can be prescribed to remove trees that are too tall and to promote growth of trees that will give the desired amount of shading.

Surveying for pests—Western spruce budworm is a serious pest of Douglas-fir in the Western United States. Timber stands are surveyed during treatment projects to determine the life stage of insects, because treatment is effective at a certain period in the life cycle of the insect. The insect reaches high populations only in certain timber types. A GIS with timber type information can be used to create a habitat map for western spruce budworm. The map is used to guide surveyors only to those areas with the potential of having high budworm populations that might need treatment. Fewer surveyors are needed, and they spend less time doing budworm surveys because they know where to survey.

Protecting species—The red-cockaded woodpecker is an endangered bird that nests in live pine trees in the Southeastern United States. Buffer zones can be placed around nest trees located in the GIS. The timber volume removed from the annual allowable cut for protection of the woodpecker can be calculated for buffer zones of several widths. This analysis relatively rapidly gives the forest supervisor information on the impact of removing nesting habitat from the resource.

Mapping pest defoliation—Forest Pest Management (FPM) personnel of the Forest Service embarked on a new and different approach to speed the process of getting gypsy moth defoliation data to the State and counties of Virginia in 1992. Forest

Service aerial photographers acquired 9-by-9-inch (22.9-by-22.9-cm) CIR stereo vertical aerial photography of Virginia using FPM force account photo aircraft and a Wild RC/10 camera at 1:50,000 scale. Forest Pest Management GIS personnel generated base maps from 1:100,000 DLG data of roads and streams for Virginia. They plotted each 1:100,000-scale data file as four 1:50,000-scale maps on transparent Mylar media. The road and stream networks were easily recognizable on the photos, which could therefore be accurately matched with the network on the Mylars. Aerial photos were cut into flight lines and placed side by side on a light table under Mylars of the same scale. Defoliation visible on the photographs was traced as polygons directly on the Mylars. The defoliation could be quickly edge-matched from both the forward lap and side lap of the aerial photos. Edge-matching is typically a major deficiency of sketch mapping because the polygons on adjoining maps typically don't match, nor do they match on State or county boundaries. Forty-four 1:50,000-scale Mylars were completed.

Geometronics personnel at the regional Forest Service office in Atlanta, Georgia, digitized the Mylars. Forest Pest Management personnel created maps of 1992 gypsy moth defoliation in Virginia from the data and sent them to Federal, State, and county cooperators. Previously, FPM personnel had hand-drawn defoliation data from media of various scales on more than 200 7.5-minute (1:24,000) USGS quadrangles. The media included (1) high-altitude U2 aircraft, with panoramic photography varying from 1:30,000 at the photo center to 1:60,000 at the edge of the photo, (2) 9-by-9-inch aerial photographs at various scales, and (3) hand-drawn sketch maps. This "eyeball" transfer process was time-consuming and error prone.

Forest Pest Management estimated a 50 percent savings in time and money using the new technique. Fewer personnel were needed to transfer data from the aerial photographs to the Mylars, and edge-matching and digitizing were much faster from the 44 map sheets used under the new method than from the 200-plus map sheets under the old method. The new method significantly increased the accuracy of data transfer by capturing data on photographs and directly transferring them to maps of the same scale.

Inventory Plot Data, Data Bases, and Reports

Data from existing inventories, data bases, and reports can be incorporated into a GIS. Plot data may be directly transferred into a GIS, if the data of the inventory are recent. More likely is the use of existing plot data to allocate samples for an updating procedure that leads to incorporation into the GIS. Other data bases may exist that allow ready incorporation or the orderly apportionment of samples. Finally, reports may contain data that can be assimilated in the new GIS. There may be a natural hierarchy to these categories, with inventories being most easily and reports least easily assimilated.

Determining coefficients of variation for future sampling—In the environment in which most GIS's are currently being constructed, it is probable that some sampling will have to be done for one resource or another. Inventory data can be used directly, but they may also be used to determine the number of samples required to achieve a given level of accuracy. Freese (1967) gives a simple and direct treatment of determining the number of samples for a given level of reliability.

$$n = \frac{t^2 s^2}{E^2} \tag{1}$$

where n is the number of samples required, t is obtained from a student's t-table (and is dependent on n), s is the standard error of the mean we are interested in estimating, and E is the error we are willing to tolerate. Remember that the t-value depends on the sample size, and n may need to be calculated iteratively.

Each of the three categories of information may be used to determine the number of required samples. Inventory plots may be used directly, and data bases may be used if the originators included estimates of sampling error, $s_{\bar{x}}$. It is unlikely that reports will have sufficient detail to allow the calculation of a sample size, unless some access to the original data is still possible.

Model-based sampling—Model-based sampling in the simplest cases is based on ratio or regression models of the relationship between a variable of interest and some other variable that is easily (inexpensively) measured. For this reason, it is important in GIS sampling.

The foundation of sampling has been the randomization principle (Hansen and others 1983). It has the function of allowing us to calculate confidence intervals from a set of input data. However, there has been a brisk discussion in the statistical literature on the use of model-based sampling. Suppose, for example, that you are interested in estimating volume of trees. Traditional sampling theory would suggest that random samples of trees be selected in order to estimate without bias the error terms and hence to derive confidence bounds on the estimate. Model-based sampling suggests that if a relationship is known to exist, it is more efficient and in some cases more accurate to use the model for selecting the sample. In forestry, we are quite accustomed to selecting sample trees to fill in certain diameter (dbh) or basal area (BA) ranges. This presumes that there is a strong relationship between dbh or BA and volume, which we know to be true. Then the error terms are estimated from the ratio or regression statistics for volume on dbh, or BA. There is considerable statistical literature on model-based sampling (Hansen and others 1983; Royall and Cumberland 1981a and 1981b), including a number of examples in forestry (such as VanDeusen 1987).

Importance sampling—Importance sampling (Rubinstein 1981) is a technique of statistical (Monte Carlo) integration that can be thought of as the continuous analogue to probability proportional to size sampling (PPS). Importance sampling has been used in forestry literature to estimate volume, weight, nutrient content, and volume growth of tree boles (Gregoire and others 1986a and 1986b, Valentine and others 1984 and 1986). Another possible set of applications could be analogous to the 3-P methods and programs of Grosenbaugh, as they were related to PPS sampling. One requisite of PPS sampling is a list of sample units prior to performing an inventory. For many forestry applications, it is unlikely that we will have a list of sample units required for PPS sampling simply because of the size of our populations. Therefore, it may be realistic to assume that we have a nearly infinite and continuous population. Until now, importance sampling has had relatively narrow application in forestry. It could be a method for obtaining point and interval estimates in cases where continuous or near-continuous variables are encountered.

Importance sampling can be simulated by constructing a proxy function for the density of the variable of interest, V. Representing the density by an integral of the form,

$$V(\lambda) = \int S(L)dL \tag{m}$$

A uniform random number is used to select a sampling point, and the actual variable S (say diameter), the proxy function for density, is measured at that point, L.

The volume is then estimated at the point by solving,

$$v_i = V(L)S(\theta_i) / V(\theta_i) \tag{n}$$

Finally, an estimate of the variable of interest for the entire tree is obtained by averaging several of these individual estimates. Point estimate and variance estimate formulas are presented in the references cited. This method is very efficient. However, it will almost always require the involvement of a statistician familiar with it.

Bayes estimation for updating—Most knowledge is based on the acquisition of information over time. Seldom do we face perfect knowledge decisions where an eternal solution can be formulated. Most foresters recognize that change is a part of management strategies. Market forces may change the value of timber that has been carefully groomed over decades so that the whole enterprise loses value. But it remains relatively unknown to foresters in general that Bayesian methods for updating, monitoring, and predicting systems exist.

Bayes models are mathematical expressions of common sense learning. So as not to confuse inflexible common sense action with common sense learning, we might say that the latter is learning from experience and adjusting to additional information in an appropriate manner.

The use of Bayes models for forecasting and updating has a relatively recent history. Still, the importance and quality of its success have made an impression on management in a number of industries. Bayes methods offer a comprehensive system for incorporating routine learning into a system to update the responsiveness of the system. The mathematical specification of the simplest model is:

$$p(Y,M) = c\,p(Y|M)p(M) \tag{o}$$

where c is a proportionality constant, usually provided by the normalizing quantity $1/p(Y)$; the model is often presented with this term instead of with the proportionality constant. In words, the model may be expressed as, "The posterior density is equal to a constant times the observed likelihood times the prior density."

To make the model more explicit, let us define D_0 as the information we have about a system initially. D_t is the information gained from some experience or sampling of the situation. Information updating can be expressed as:

$$D_t = \{I_t, D_{t-1}\},\ (t = 1, 2, \ldots). \tag{p}$$

The time (or spatial) sequence for obtaining statistics to project into a future prediction of $Y_t, Y_{t+1}, \ldots$, is conditioned on D_t. These statistical models are dependent on parameters of statistical models, such as means, variances, and forms of distributions, which we represent in vector form (bold type) as $\mathbf{q}_t$. Then the one-step-ahead prediction with a parameter-dependent model is:

$$p(Y_t | \theta_t, D_{t-1}). \qquad (q)$$

The parameters summarized by $\mathbf{q}_t$ must represent meaningful summaries for the forecasting problem. Usually, the parameter set will be fixed, though the values may change in dynamic systems. In some cases, especially where social, political, or other model shifts occur, the parameter set may change over time! The model parameters $\mathbf{q}_t$ are the means by which information about the process are incorporated in the model, and the learning process involves sequentially revising the state of knowledge about these parameters.

There are some important considerations that need to be adequately addressed before Bayes methods can be applied. First, the sequence of events needs to be temporally or spatially equidistant. The difficulty of applying the models to samples from random times is not always insuperable, but always requires much more effort. Second, the variance-covariance structure of the overall model has a pronounced effect on the resulting predictions or updates. It is not a simple task to obtain good estimates for the variance-covariance. There are two components: the error associated with the mean of observations, and the variance associated with the system. Lack of knowledge about the distribution of these two components can lead to poor updating and predictions. Simply estimating them from the first set of data will not do! Then, too, the number of intervals (remember, these may be in time or distance) in the model have an importance that is sometimes ignored. Very short time series may be modeled in the Bayes framework, but the estimates will be very little better than repeated composite estimation until the structure of the system and observation error are well supported by adequate data.

For example, let us suppose the forest needs to model expected price for thinnings from a series of overstocked stands. For some time, the demand for pulpwood has been relatively constant, and hence monthly mean price has been relatively stable (constant average) with minor variation between months. (The Bayes procedure can also model interventions or shocks, but we will not deal with those here.) Suppose that data from a large sample of stands on the Enchanted Forest is pooled to obtain initial estimates. The price is $13 per cord, and the variation is first estimated from experience to be plus or minus about $4, and, from rules of thumb, the variance is 40. So

$$(\mu_0 | D_0) \sim N(13, 40)$$

From this distribution, one can obtain:

$$Y_t = \mu_t + v_t, \quad v_t \sim N(0,10)$$

and

$$\mu_t = \mu_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim N(0,.5)$$

which represents a .05 signal-to-noise ratio. Observations and some components are given in table 10.

Table 10—Bayes updating of monthly pulpwood prices.

Month	Forecast distribution		Observation	Error	Posterior information	
t	Q_t	f_t	Y_t	e_t	m_t	C_t
0	—	—	—	—	13	40
1	50	13	15	2	14.6	8
2	18	14.6	13.6	−1	14.1	4.6
3	15	14.1	14.3	.2	14.2	3.4
4	14	14.2	15.4	1.2	14.5	2.8
5	13	14.5	13.5	−1	14.3	2.5
6	13	14.3	14.8	.5	14.4	2.3
7	13	14.4	12.8	−1.6	14.0	2.2
8	12.5	14.0	14.9	.9	14.2	2.1
9	12.5	14.2	14.6	.4	14.3	2.0
10	12.5	14.3	—	—	—	—

Not shown in the example is the computed weighting factor, which, like the composite estimator, is based on the variance. As the number of periods increases, the size of the weight function decreases rapidly. Note also that by $t = 10$, this weighting would mean that m_0 contributes only 1 percent of the information used to calculate m_{10}. We repeat: information that is not directly quantifiable can often be incorporated into a Bayes updating procedure. If, for instance, mills began introducing hardwood fiber in their furnish, a reasonable assumption would be that the price of softwood would fall, and it could easily be modeled as an intervention.

Bayes methods have been increasingly applied to problems in prediction for things as disparate as the stock market and positions of space ships. They can have an increasingly important role in "learning" about complex functional relationships in ecosystems; however, they almost always require significant input of information. They especially require information about the variance and covariance of data.

Foresters will have to finally begin to pay attention to data collection that allows the calculation of these variance components if they wish to benefit from the advantages these computational schemes allow.

Summary

Incorporation of existing information in the establishment of a GIS is a critical part of the process. The care and maintenance of information can make or break the utility of a GIS. Our credibility as managers will increasingly rest on the quality of the information about our resources as they are distributed over the land base. A variety of statistical and other quantitative methods that may have seemed too complex to be useful in natural resource management have suddenly become feasible tools for the establishment, care, and maintenance of our basic data. Existing data can be prepared for inclusion in corporate data bases and GIS's and for designing future resource inventories through the appropriate use of remotely sensed data as well as available photography and its associated manipulation. Techniques available for utilizing existing information include data conversion, updating and maintaining information, and the use of data for improving new mapping and inventory projects. We have tried to provide an idea of the statistical techniques, Bayes estimation, model-based sampling, and composite estimation that are now available for application to the problems that land managers face and that a GIS may provide solutions to in the future. The use of available information saves time and money, and provides links to the activities of other users both within and among natural resource agencies and other organizations.

Chapter 7: Building Better Data Bases

This primer provides general guidance on determining resource information needs, locating information, evaluating information for use in corporate data bases, and using it in the decisionmaking process. We have attempted to cover the most salient requirements of constructing and maintaining natural resource-based information systems. Many emerging technologies will influence the choice of techniques for computation and analyses. We hope our coverage is broad enough to encourage new users as well as deep enough in the matter of mathematical and technological developments to point more advanced users in the right direction so that at least some pitfalls might be avoided. Practitioners and those needing more detailed direction are encouraged to consult the literature listed at the end of this report.

A successful data base is one that provides the principal users and stakeholders with the environmental, economic, and social information they need to make sound and timely decisions, and to understand the predicted risks of alternative decisions; the format is one the principal user can understand and manipulate. The ideal data base contains information that the decisionmaker needs for a variety of problems—without redundant data, but also without gaps. Attempting to provide data that will meet tomorrow's needs as well as today's is a challenge that requires our best efforts at anticipating management needs. Preparing today's data base for tomorrow's needs includes updating existing information and incorporating auxiliary information. Chapter 2 outlines procedures for identifying information requirements.

Managers would like to understand the range of projected outcomes of their decisions. Figure 16 shows the sources of information available to the decisionmaker.

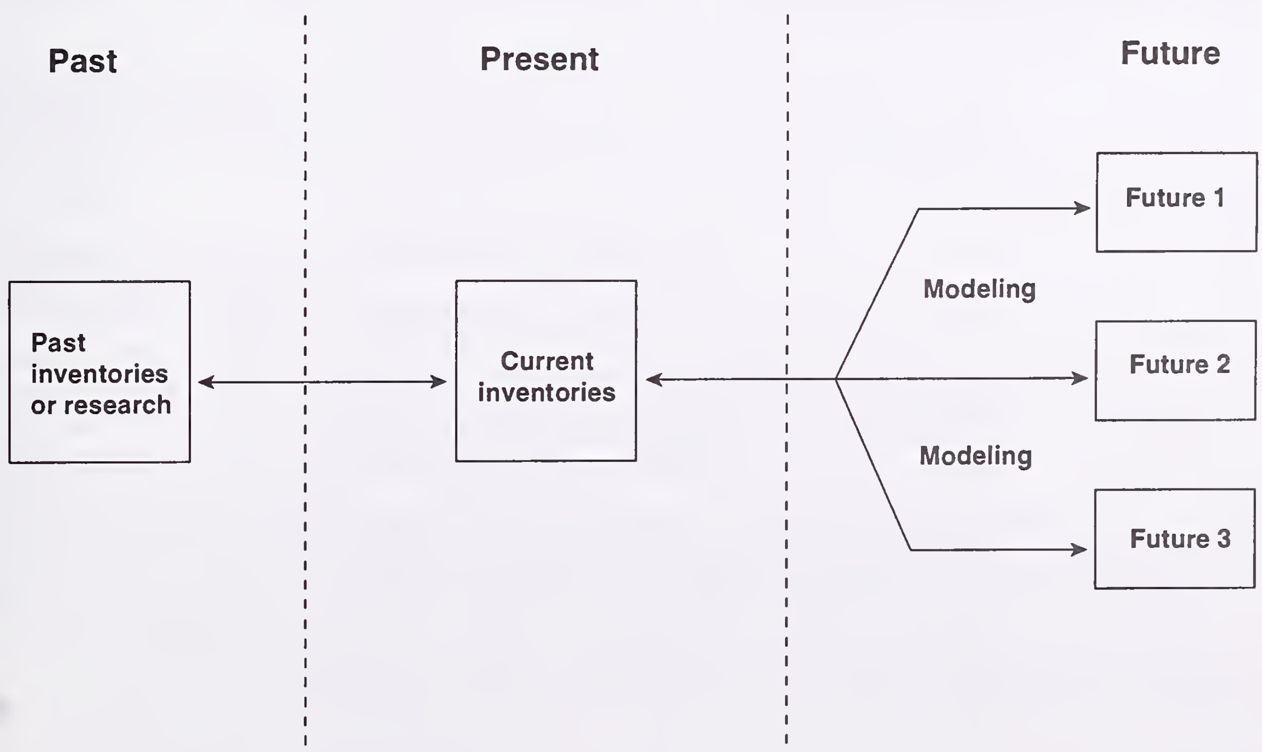


Figure 16—Sources and flow of information for decisionmakers.

Current inventories provide the resource base for all decisions. Resource inventories populate data bases that are, in turn, used to generate assessments. Assessments, which include outlooks for the future, are then hopefully used for sustainable development. To determine the outcome of future decisions, the resource specialist must know how resources have responded to past treatments. Using experience gained, analysts model the future based upon the present situation and alternative assumptions.

As chapter 3 shows, information is abundant, diverse, and increasing daily. New satellites are being planned and launched. Heightened environmental awareness has expanded the pool of individuals and organizations collecting data, from environmental groups to industries, from students to international agencies. At the same time, our understanding of our environment is increasing, and new modeling tools are becoming available.

Not all information is useful for a given need. Chapter 4 discusses data utility and alerts the reader to sources of error and the kinds of errors to watch for. Data base producers, distributors, and managers are responsible for ensuring the existence and availability of appropriate data quality and documentation for any large environmental data base (Goodchild 1994). Producers, distributors, and managers are encouraged to read this chapter and accept this responsibility.

Funds and personnel for forest management have been declining in recent years. Every dollar must be spent effectively. We cannot afford to continually gather data for each function. Corporate data bases that contain a core set of widely used data for sharing information should be created. Data that are to be shared and used by others require special consideration. We need to ensure that these users are not receiving faulty information or information that may be misunderstood or misinterpreted. Chapter 5 describes ways to evaluate existing information from a corporate and GIS standpoint.

Not all existing data will meet our information needs. Eventually, new data must be collected. But some existing information can assist us in reducing inventory costs, and chapter 6 shows how.

In the movie classic "The Wizard of Oz," a Kansas tornado transported Dorothy from a bleak and barren black and white landscape to a beautifully colored world. Computer technology allows us to do something similar. With personal computers, image analysis systems, and GIS's, bits of black and white information can be translated into very vivid and colorful displays. Dorothy's Land of Oz was a fantasy, of course. As resource and information specialists, we need to ensure that the world we create for our managers and decisionmakers is rooted in reality. There is an old computer adage—"Garbage in, garbage out," or "GIGO" for short. Good information is essential for good resource management. We hope that the guidance given in this primer will lead to the creation of globally better data bases.

References

Aldred, A.H. 1981. A federal/provincial program to implement computer-assisted forest mapping for inventory and updating. Ottawa, ON: Dendron Resource Surveys Ltd. 81 p.

American Society of Civil Engineers. 1978. Definitions of surveying and associated terms. Washington, DC: American Society of Civil Engineers and American Congress on Surveying and Mapping. 210 p.

Anonymous. 1991. Data acquisition guidelines for geographical information systems. Missoula, MT: U.S. Department of Agriculture, Forest Service, Northern Region. 53 p.

Aronoff, S. 1989. Geographic information systems: a management perspective. Ottawa, ON: WDL Publications. 294 p.

Bailey, R.G. 1988a. Ecogeographic analysis: a guide to the ecological division of land for resource management. Misc. Publ. 1465. Washington, DC: U.S. Department of Agriculture, Forest Service. 16 p.

Bailey, R.G. 1988b. Problems with using overlay mapping for planning and their implications for geographic information systems. Environmental Management 12(1):11–17.

Bain, S. 1991. Personal correspondence. USDA Forest Service, Washington, DC.

Battese, G.E.; Harter, R.M.; Fuller, W.A. 1988. An error-component model for prediction of county crop areas using survey and satellite data. Journal of the American Statistical Association 83:28–36.

Bauer, M.E.; Hixson, M.M.; Davis, B.J.; Etherridge, J.B. 1978. Area estimation of crops by digital analysis of Landsat data. Photogrammetric Engineering and Remote Sensing 44:1033–1043.

Berry, J.K. 1989. Introduction to geographic information systems (GIS) and computer-assisted map analysis. Management, mapping, and analysis for geographic information systems workshop workbook. Fort Collins, CO: Berry & Associates. Misc. pagination.

Burrough, P.A. 1986. Principles of geographical information systems for land resources assessment. New York: Oxford University Press. 194 p.

Burrough, P.A. 1989. Matching spatial databases and quantitative models in land resource assessment. Soil Use and Management 5(1):3–8.

Calkins, H.W. 1983. A pragmatic approach to geographic information system design. In: Pequet, D.; O'Callaghan, J., eds. Design and implementation of computer-based geographic information systems. Australian Workshop. Honolulu, HI: IGU Commission on Geographical Data and Processing.

Card, D.H. 1982. Using known map category marginal frequencies to improve estimates of thematic map accuracy. Photogrammetric Engineering and Remote Sensing 48(3):431–439.

Chen, J. [and others]. 1991. ZOMAP—a knowledge-based geographical information system for zoning. In: GIS/LIS Proceedings; 1991 October 28–November 1; Atlanta, GA. Washington, DC: American Congress of Surveying and Mapping: 150–155.

Chhikara, R.S.; Lundgren, J.C.; Houston, A.A. 1986. Crop acreage estimation using a Landsat-based estimator as an auxiliary variable. International Transactions in Geoscience and Remote Sensing 24:157–168.

Chhikara, R.S; Deng, L.Y. 1988. Conditional inference in finite population sampling under a calibration model. Communications in Statistics—Simulation and Computation 17:663–681.

Chrisman, N.R. 1982. Beyond accuracy assessment: correction of misclassification. In: Foreman, J., ed. Fifth International Symposium on Computer-assisted Cartography and International Society for Photogrammetry Remote Sensing Commission IV: Cartographic and Data Bank Application of Photogrammetry and Remote Sensing; 1982 August 22–28; Crystal City, VA. Falls Church, VA: American Society for Photogrammetry and Remote Sensing: 123–132.

Colwell, R.N., ed. 1983. Manual of remote sensing. 2d ed. Falls Church, VA: American Society for Photogrammetry. 2,440 p.

Czaplewski, R.L. 1990a. Combining inventories of land cover and forest resources with prediction models and remotely sensed data. In: Lund, H. G.; Preto, G., eds. Global natural resource monitoring and assessments: preparing for the 21st Century. Proceedings of an international conference; 1989 September 24–30; Venice, Italy. Bethesda, MD: American Society for Photogrammetry and Remote Sensing: 1,078–1,088.

Czaplewski, R.L. 1990b. Updating land cover/land use estimates using the Kalman Filter. In: Lund, H. G.; Preto, G., eds. Global natural resource monitoring and assessments: preparing for the 21st Century. Proceedings of an international conference; 1989 September 24–30; Venice, Italy. Bethesda, MD: American Society for Photogrammetry and Remote Sensing: 589–590.

Czaplewski, R.L.; Catts, G. P. 1990. Calibrating area estimate for classification error using confusion matrices. In: Image processing/remote sensing. Vol. 4, Technical Papers, 1990 ACSM-ASPRS Annual Convention; 1990 March 18–23; Denver, CO. Techn. Papers 4. Bethesda, MD: American Congress on Surveying and Mapping and American Society for Photogrammetry and Remote Sensing: 431–440.

Czaplewski, R.L.; Catts, G. P. 1992. Calibration of remotely sensed proportional area estimates for misclassification of error. Remote Sensing of Environment 39:29–43.

Department of Defense. 1981. Glossary of mapping, charting, and geodetic terms. 4th ed. Washington, DC: U.S. Department of Defense, Defense Mapping Agency. 204 p.

Dull, C.W.; Mead, R.; Yockey, E.; Clerke, B.; Ashby, J.; Johnson, R.; Heins, R. 1989. Geographic information system implementation plan, Region 8. Atlanta, GA: U.S. Department of Agriculture, Forest Service, Southern Region. 17 p.

Evanisko, F. 1990. Ecosystem mapping: possibilities and impediments for remote sensing. In: Greer, J. D., ed. Proceedings, Third Biennial, USDA Forest Service Remote Sensing Applications Conference; 1990 April 8–13; Tucson, AZ. Bethesda, MD: American Society for Photogrammetry and Remote Sensing: 403–413.

Federal Geographic Data Committee. 1993. Manual of federal geographic data products. Falls Church, VA: ViGYAN Inc. Misc. pagination.

Ford-Robertson, F.C., ed. 1971. Terminology of forest science, technology, practice, and products. Multilingual Forestry Terminology Series No. 1. Washington, DC: Society of American Foresters. 349 p.

Freedman, A. 1983. The computer glossary for everyone: it's not just a glossary. Englewood Cliffs, NY: Prentice-Hall, Inc. 320 p.

Freese, F. 1967. Elementary statistical methods for foresters. Agriculture Handbook 317. Washington, DC: U.S. Department of Agriculture, Forest Service. 87 p.

Goodchild, M., chair, 1994. Developing large environmental data bases for sustainable development. In: Singh, A., ed. Proceedings, UNEP/IUFRO International Workshop in Cooperation with FAO on Developing Large Environmental Data Bases for Sustainable Development; 1993 July 14-16; Nairobi, Kenya. GRID Inf. Ser. No. 22. Nairobi, Kenya: United Nations Environment Programme: 356–368.

Goodchild, M.F.; Gopal, S., eds. 1989. Accuracy of spatial data bases. Bristol, PA: Taylor and Francis, Ltd. 290 p.

Goodchild, M.F.; Wang, M.H. 1988. Modeling error in raster-based spatial data. In: Christman, N. R., ed. Proceedings, Third International Symposium of Spatial Data Handling; 1988 August; Sydney, Australia. Columbus, OH: International Geographical Union, Commission on Geographical Data Sensing and Processing: 97–106.

Grassia, A.; Sundberg, R. 1982. Statistical precision in the calibration and use of sorting machines and other classifiers. Technometrics 24:117–121.

Green, K.; Congalton, R. G. 1991. Mapping potential old growth forests and other resources on national forests and park lands in Oregon and Washington. In: GIS/LIS Proceedings; 1990 November 7–10; Anaheim, CA. Bethesda, MD: American Society for Photogrammetry and Remote Sensing: 712–723.

Green, E.J.; Strawderman, W.E. 1986. Reducing sample size through the use of a composite estimator: an application to timber volume estimation. Canadian Journal of Forest Research 16(5):1116–1118.

Gregoire, T.G.; Valentine, H.T.; Furnival, G.M. 1986a. Estimation of volume with a taper function: A sampling approach. In: Oderwald, R. G.; Burkhart, H. E.; Burk, T. E., eds. Use of auxiliary information in natural resource inventories. Proceedings; 1985 October 1–2; Blacksburg, VA. SAF 86-01. Blacksburg, VA: Virginia Polytechnic Institute and State University, School of Forestry and Wildlife Resources: 164–170.

Gregoire, T.G.; Valentine, H.T.; Furnival, G.M. 1986b. Estimation of bole volume by importance sampling. Canadian Journal of Forest Research 16:554–557.

Gregoire, T.G.; Walters, D.K. 1988. Composite vector estimators derived by weighting inversely proportional to variance. Canadian Journal of Forest Research 18(2): 282–284.

Haddon, B.D. 1988. Forest inventory terms in Canada. Ottawa, ON: Forestry Canada, Canadian Forest Inventory Committee. 109 p.

Hansen, M.H.; Madow, W.G.; Tepping, B.J. 1983. An evaluation of model-dependent and probability-sampling inferences in sample surveys. Journal of the American Statistical Association 78(384):776–807.

Hay, A.M. 1988. The derivation of global estimates from a confusion matrix. International Journal of Remote Sensing 9:1395–1398.

Heydorn, R.P.; Takacs, H.C. 1986. On the design of classifiers for crop inventories. IEEE Transactions in Geoscience and Remote Sensing 24:150–155.

Hoekstra, T.W. 1982. Information needs and use assessment. In: Brann, T.B.; House, L.O. IV; Lund, H.G., eds. In-place resource inventories: principles and practices. Proceedings, National Workshop, Society of American Foresters Inventory Working Group; 1981 August 9–14; Orono, ME. SAF 82-02. Bethesda, MD: Society of American Foresters: 10–13.

Hung, H.M.; Fuller, W.A. 1987. Regression estimation of crop acreages with transformed Landsat data as auxiliary variables. Journal of Business and Economic Statistics 5: 475–482.

Iverson, L.R.; Cook, E.A.; Graham, R.L. 1989. A technique for extrapolating and validating forest cover across large regions, calibrating AVHRR data with TM data. International Journal of Remote Sensing 10:1805–1812.

Kendall, M.G.; Buckland, W.R. 1971. A dictionary of statistical terms. 3d. ed. New York: Hafner Publishing Co. 166 p.

Kish, L. 1965. Survey sampling. New York: John Wiley & Sons. 621 p.

Lachowski, H. 1990. Report to Remote Sensing Steering Committee. Remote sensing working group notes. Unpublished report. Salt Lake City, UT: U.S. Department of Agriculture, Forest Service, Nationwide Forestry Applications Program. 20 p.

Landrum, V.; Lachowski, H.; Steffenson, J. 1991. Data storage needs for remotely sensed digital data. Report—U.S. Department of Agriculture, Forest Service, Nationwide Forestry Applications Program. Salt Lake City, UT. 4 p.

Langley, P.G., 1975. Multistage variable probability sampling: theory and use in estimating timber resources from space and aerial photography. Berkeley, CA: Univ. of California. Publication number 75-22538. 101p. Dissertation.

Langley, P.G.; von Roessel, C.S.; Place, M.C. 1980. Multiresource inventory methods pilot test (Phase I): evaluation of multiresource analysis and information system (MAIS) processing components, Kershaw County, SC, feasibility test. U.S. Department of Agriculture, Forest Service, AgRISTARS, Renewable Resource Inventory Project, Earth Satellite Corporation Contract Report 80NFAP-211. Houston, TX: U.S. Department of Agriculture, Forest Service. 96p.

Lewis, T.O.; Odell, P.L. 1971. Estimation in linear models. Englewood Cliffs, NJ: Prentice-Hall. 193 p.

Liew, C. K. 1976. Inequality constrained least-squares estimation. Journal of the American Statistical Association 71:746–751.

Lund, H.G. 1985. Designing inventories to support multiple decisions. Unpublished paper presented at Forest Resource Inventories: Appraisal Tools for Decision Making; 1985 April 26–28; Venice, Italy.

Lund, H.G. 1986a. A primer on integrating resource inventories. Gen. Tech. Rep. WO-49. Washington, DC: U.S. Department of Agriculture, Forest Service. 64 p.

Lund, H.G. 1986b. Existing information and the sorcerer's apprentice. In: Oderwald, R.G.; Burkhart, H.E.; Burk, T.E., eds. Proceedings, Use of Auxiliary Information in Natural Resource Inventories; 1985 October 1–2; Blacksburg, VA. SAF 86-01. Blacksburg, VA: Virginia Polytechnic Institute and State University, School of Forestry and Wildlife Resources: 1–8.

Lund, H.G. 1987. Developing resource inventory policies for national land and resource evaluation planning. In: Lund, H.G.; Caballero-Deloya, M.; Villareal-Canton, R., eds. Land and Resource Evaluation for National Planning in the Tropics, Proceedings, International Conference and Workshop; 1987 January 25–31; Chetumal, Mexico. Gen. Tech. Rep. WO-39. Washington, DC: U.S. Department of Agriculture, Forest Service: 491–498.

Lund, H.G. 1990. In search of blackbirds, big brothers, bull puckey, and the unknown: issues for working group discussions. In: Greer, J.D., ed. Proceedings, Third Biennial, USDA Forest Service Remote Sensing Applications Conference; 1990 April 8–13; Tucson, AZ. Bethesda, MD: American Society for Photogrammetry and Remote Sensing: 80–85.

Lund, H.G.; Mohil El Deen, F.A.; Allison, R.P., Jasumback, T. 1990. The Sudan reforestation and anti-desertification (SRAAD) project: applications for watershed management planning. In: Potter, C.F., ed. Proceedings, Planning for Sustainable Watershed Management: Environmental and Institutional Assessments, Interdisciplinary Workshop; 26–27 June 1990; Washington, DC. Washington, DC: Agency for International Development, Bureau for Science and Technology: 84–98.

Lund, H.G.; Schreuder, H.T. 1980. Aggregating inventories. Res. Eval. Newsl. 4. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Forest and Range Experiment Station: 1–3.

Lund, H.G.; Thomas, C.E. 1989. A primer on stand and forest inventory designs. Gen. Tech. Rep. WO-54. Washington, DC: U.S. Department of Agriculture, Forest Service. 96 p.

Marriott, F. H. C. 1990. A dictionary of statistical terms. 5th ed. Essex, England: Longman Scientific and Technical Press. 223 p.

Maxim, L.D.; Harrington, L.; Kennedy, M. 1981. Alternative scale-up estimates for aerial surveys where both detection and classification error exist. Photogrammetric Engineering and Remote Sensing 47:1227–1239.

Mereszczak, I., Colmer, J.; Oden, J.; Almy, J.; Hoppus, M.; Ishikawa, P., Jr.; Lachowski, H.; Ward, J.; Oslin, A. J. 1990. Riparian area management: what can remote sensing contribute? EM 7140-19. Washington, DC: U.S. Department of Agriculture, Forest Service, Engineering Staff. 70 p.

Odum, E.P. 1971. Fundamentals of ecology. Philadelphia, PA: W.B. Saunders Co. 574 p.

Ogrosky, C. 1992. Briefing on redraft of proposed U.S. cartographic standards for spatial accuracy. Report—Federal Geographic Data Committee, Base Cartographic Data Subcomittee. Washington, DC. 3 p.

Pech, R.P.; Davis, A.W.; Lamacraft, R.R.; Graetz, R.D. 1986. Calibration of Landsat data for sparsely vegetated semi-arid rangelands. International Journal of Remote Sensing 7:1729–1750.

Poso, S.; Paananen, R.; Simila, M. 1987. Forest inventory by compartments using satellite imagery. Silva Fennica 21(1):69–94.

Prisley, S.P.; Smith, J.L. 1987. Using classification error matrices to improve the accuracy of weighted land-cover models. Photogrammetric Engineering and Remote Sensing 53:1259–1263.

Richards, J.A. 1986. Remote sensing digital image analysis. New York: Springer-Verlag. 281 p.

Rosenfeld, C.; Cooke, R. 1982. Earthfire—the eruption of Mount St. Helens. Cambridge, MA: The MIT Press. 155 p.

Rosson, J.F.; McWilliams, W.H.; Frey, P.D. 1988. Forest resources of Louisiana. Res. Bull. SO-130. New Orleans, LA: U.S. Department of Agriculture, Forest Service, Southern Forest Experiment Station. 81 p.

Royall, R.M.; Cumberland, W. G. 1981a. An empirical study of the ratio estimator and estimators of its variance. Journal of the American Statistical Association 76(373): 66–88.

Royall, R.M.; Cumberland, W. G. 1981b. The finite population linear regression estimator and estimators of its variance—an empirical study. Journal of the American Statistical Association 76(376): 924–930.

Rubinstein, R. Y. 1981. Simulation and the Monte Carlo method. New York: John Wiley & Sons. 278 p.

Schaible, W. L. 1978. Choosing weights for composite estimators for small area statistics. In: Proceedings of the American Statistical Association, Sec. on Survey Res. Methods: 741–746.

Shim, J.K. 1983. A survey of quadratic programming applications to business and economics. International Journal of Systems Science 14:105–115.

Siegal, S.; Castellan, N.J., Jr. 1988. Nonparametric statistics for the behavioral sciences. 2nd edition. McGraw-Hill Co., New York, NY. 399 p.

Slama, C.C., ed. 1980. Manual of photogrammetry, 4th ed. Falls Church, VA: American Society for Photogrammetry and Remote Sensing. 1,056 p.

Sorenson, H.W. 1985. Kalman filtering: theory and applications. New York: IEEE Press Inc. 457 p.

Spatial Data Transfer Committee. 1979. Standard format for the transfer of geocoded polygon data. Ottawa, ON.: Energy, Mines and Resources. 71 p.

Spears, B.M.; Nettleton, W.A. 1993. Use of a geographic information system in southern pine beetle management. In: Liehbold, A.M.; Barrett, H.R., eds. Proceedings, Spatial Analysis of Forest Pests; 1992 April 27–30; Mountain Lakes, VA. Gen. Tech. Rep. NE-175. Radnor, PA: U.S. Department of Agriculture, Forest Service, Northeastern Forest Experiment Station: 15–24.

Swain, P.H.; Davis, S.M. 1978. Remote sensing: the quantitative approach. New York: McGraw-Hill, Inc. 396 p.

Tenenbein, A. 1972. A double sampling scheme for estimating from misclassified multinomial data with applications to sampling inspection. Technometrics 14:187–202.

Teply, J. 1991. Personal correspondence. USDA Forest Service, Portland, OR.

Texas A&M University. 1992. INFORMS-TX project review. College Station, TX: Department of Rangeland Ecology and Management, STARR Laboratory. 40 p.

Thomas, C.E.; Rennie, J.C. 1987. Combining inventory data for improved estimates of forest resources. Southern Journal of Applied Forestry 11(3):168–171.

U.S. Department of Agriculture, Forest Service. 1988a. National GIS plan. Washington, DC: U.S. Department of Agriculture, Forest Service, Information Systems. 14 p.

U.S. Department of Agriculture, Forest Service. 1988b. Resource information project. Washington, DC: U.S. Department of Agriculture, Forest Service, Information Systems. 68 p.

U.S. Department of Agriculture, Forest Service. 1989. Interim resource inventory glossary. Washington, DC: U.S. Department of Agriculture, Forest Service. 96 p.

U.S. Department of Agriculture, Forest Service. 1990a. Region 6 guidelines for map manuscript preparation and digitizing (draft glossary). Portland, OR: U.S. Department of Agriculture, Forest Service, Pacific Northwest Region. Misc. pagination.

U.S. Department of Agriculture, Forest Service. 1990b. Resource inventory handbook. FSH 1909.14 WO Amendment 1, effective 1990 March 29. Washington, DC: U.S. Department of Agriculture, Forest Service. Misc. pagination.

U.S. Department of Agriculture, Forest Service. 1990c. Service-wide standard terms and definitions for resource data. 1390 letter dated 1990 December 3. Washington, DC: U.S. Department of Agriculture, Forest Service. 144 p.

U.S. Department of Agriculture, Forest Service; Department of Entomology, Texas A&M University; Department of Computer Science, Texas A&M University. 1992. ISPBEX project review. Atlanta, GA: U.S. Department of Agriculture, Forest Service, Southern Region, Forest Pest Management. 33 p.

U.S. Department of Interior Geological Survey. 1985. U.S. GeoData—digital cartographic and geographic data. Reston, VA: U.S. Department of Interior, Geological Survey, National Cartographic Information Center. 1 p.

U.S. Department of Interior Geological Survey. 1990. Digital elevation models data users guide. Reston, VA: U.S. Department of the Interior, Geological Survey. 47 p.

U.S. Department of Interior Geological Survey. 1992a. Spatial data transfer standard (FIPS 173). Reston, VA: U.S. Department of the Interior, Geological Survey, SDTS Task Force. 6 p.

U.S. Department of Interior Geological Survey. 1992b. United States national cartographic standards for spatial data (draft). Reston, VA: U.S. Department of the Interior, Geological Survey. 4 p.

Valentine, H.T.; Gregoire, T.G.; Furnival, G.M.; Solomon, D.S. 1986. Unbiased estimation of bole increment. In: Environmental influences on tree and stand increment. Misc. Publ. 691. U.S. Department of Agriculture, Forest Service, Maine Experimental Station: 67–72.

Valentine, H.T.; Gregoire, T.G.; Furnival, G.M. 1984. Subsampling trees for biomass, volume or mineral content. Forest Science. 30:673-681.

Valentine, W. 1990. Accurate data for strategic and project GIS. In: Greer, J.D. ed. Proceedings, Third Biennial, U.S. Department of Agriculture Forest Service Remote Sensing Applications Conference; 1990 April 8–13; Tucson, AZ. Bethesda, MD: American Society for Photogrammetry and Remote Sensing: 368–373.

VanDeusen, P. C. 1987. 3-P sampling and design-based estimates. Canadian Journal of Forest Research 17(2):115–117.

Veregin, H. 1989. A taxonomy of error in spatial databases. Tech. Paper 89-12 (Re/I-1). Santa Barbara, CA: University of California, Geography Department, National Center for Geographic Information and Analysis. 115 p.

Walters, C.J.; C.S. Holling. 1990. Large-scale management experiments and learning by doing. Ecology 71(6):2060–2068.

Ware, K.D.; Cunia, T. 1962. Continuous forest inventory with partial replacement. For. Sci. Monogr. No. 3. Washington, DC: Society of American Foresters. 40 p.

Warnecke, L. [and others]. 1992. The state geographic information activities compendium. Lexington, KY: The Council of State Governments. 601 p.

Winterberger, K.; Reutebuch, S. 1990. Working group C (bull puckey) recommendations related to the evaluation of existing data and imagery for input to GIS and corporate data bases. In: Greer, J.D., ed. Proceedings, Third Biennial, U.S. Department of Agriculture Forest Service Remote Sensing Applications Conference; 1990 April 8–13; Tucson, AZ. Bethesda, MD: American Society for Photogrammetry and Remote Sensing: 432–435.

Young, P. 1984. Recursive estimation and time-series analysis. New York: Springer-Verlag. 300p.

Appendix: Summary of Remote Sensing Sources

Imagery may be acquired either photographically or through electro-optical scanners.

Aerial Photography

Aerial photography is the most widely used form of remote sensing imagery in geophysical, cartographic, and natural resource management applications. Modern camera systems and films can provide imagery of high resolution over a broad range of scales. Aerial photographic systems record reflected energy in the visible and near-infrared portions of the spectrum. Factors that define the utility of aerial photography include coverage, date of mission, scale of imagery, emulsion of film, format of camera, focal length of lens, and atmospheric conditions at time of mission.

Aerial photographic imagery is available in scales ranging from 1:2,000 to 1:2,000,000 to meet a broad range of resource analysis requirements. The largest scales are used for engineering studies and site development activities. Imagery at scales from 1:12,000 to 1:24,000 are most frequently used for resource management applications. Cartographers use photography at scales from 1:24,000 to 1:80,000 for map production. Aerial photography imagery acquired by reconnaissance aircraft has proved useful in supporting a wide range of USDA Forest Service requirements for data over extensive areas at scales ranging from 1:30,000 to 1:120,000 (Hinkle 1981). Small-scale photographic coverage of extensive areas is available from both the U.S. and Russian space programs. Areas covered by individual aerial photography missions range from a few frames for development of specific sites to missions covering entire management units or States.

Aerial photography is available from private and government organizations for natural resource management, agricultural, engineering, and cartographic applications. Depending on the users' requirements, either historic or current coverage may be required. The Aerial Photography Summary Record System, maintained by the Earth Sciences Information Centers operated by the U.S. Geological Survey (USGS), indexes the holdings of over 500 cooperating Federal, State, and private organizations. Imagery may be ordered from individual agencies based on the results of computerized searches against user requirements. See Federal Geographic Data Committee report (1993) and Warnecke and others (1992) for additional information on data available from Federal and State agencies.

Electro-optical Systems

Electro-optical systems are capable of recording electromagnetic energy extending from the ultraviolet to the thermal infrared portion of the spectrum. There are two classes of electro-optical sensors: nonimaging and imaging. Nonimaging sensors acquire individual measurements rather than an array of measurements that form an image. Spectrometers mounted on aircraft or bucket trucks are used to acquire reflectance measurements from scene elements (vegetation, water bodies, etc.) to calibrate airborne and satellite imagery or develop reference data. Nonimaging

sensors, carried aboard aircraft, obtain measurement profiles for specialized applications. Examples of nonimaging sensors used in natural resource applications include airborne laser profilers and instruments such as the Airborne Oceanographic Lidar (Hoge and others 1983) that measure the laser-induced florescence of ground features.

These electro-optical systems are not limited by the sensitivity of chemical reactions that occur when reflected light strikes the film in an aerial camera to create an image. Information from electro-optical systems may be recorded in analog or digital format. Video systems and some radar systems capture data in analog form, but most electro-optical systems convert the intensity of incoming energy directly to digital data. Although generally of lower spatial resolution than aerial photography, electro-optical sensor data have advantages for natural resource applications. Image analysts can directly manipulate the digital imagery using computer-based systems to rectify, classify, enhance, and display the imagery. Electro-optical sensors capture information beyond the visible spectrum.

Electro-optical systems can be configured to acquire information from the ultraviolet through the visible, near, middle, and thermal infrared to the microwave portion of the spectrum. The middle and thermal infrared portions of the spectrum are important in identifying and assessing the condition of vegetation. Radar systems can acquire imagery through cloud cover and at night. Satellite imagery and airborne scanners are the electro-optical systems most widely used in natural resource management. Airborne video systems are becoming an important tool for acquiring data rapidly at a low cost for limited areas.

Remote sensors carried aboard earth-orbiting satellites provide digital imagery of extensive areas. Earth resources satellite systems provide imagery with a resolution suitable for many natural resource management requirements (30 to 98 feet or 10 to 30 meters). These systems provide imagery at a low cost per acre, with consistent mission parameters and repetitive coverage. The repeat coverage cycle of earth resources satellites now in orbit is from 14 to 16 days.

Airborne Sensors

Airborne video is a relatively recent addition to the spectrum of remote sensing systems available for natural resource applications. Video systems have lower resolutions than comparable photographic systems and currently lack calibration necessary for precision photogrammetric applications. They are well suited for many natural resource applications requiring sample or small area coverage. They are also cost-effective for locating features such as isolated groups of insect-damaged trees within a larger survey area. System operators can evaluate video data during acquisition and change mission parameters as necessary. Improvements in camera design and the advent of higher definition recording formats such as Super VHS and HI-8 video have increased the resolution and utility of video systems for natural resource application. Image analysts can manually interpret video imagery using a high-resolution monitor and a playback unit with freeze-frame capability. For enhancement and georeferencing, the analog data in individual video frames can be captured as digital data using a video "frame-grabber." The relatively low cost of video systems makes them a good candidate for many monitoring applications (Myhre and others 1991).

Airborne electro-optical remote-sensing systems cover a broad range of capabilities. Airborne systems support working requirements and serve as test beds to evaluate new sensor designs. Contractors, research and development organizations, and State and Federal agencies operate airborne, electro-optical remote sensing systems. The Forest Service operates airborne thermal infrared systems to support fire suppression activities. The National Aeronautics and Space Administration (NASA) designs sensor systems and conducts application research. Description of sensor capabilities and information on available data can be obtained from the NASA centers with aircraft programs. It is possible to obtain existing data from almost any of these organizations. One must recognize that the data from many current airborne digital remote sensing systems are difficult and expensive to register to ground coordinates. In addition, specialized software and knowledge may be necessary to extract useful information from these data. Nevertheless, airborne systems have an extremely wide range of capabilities and the potential for providing solutions to many unique requirements.

Radar data, although more difficult to process and with lower resolution, have the advantage of operating through clouds or at night.

Satellite Systems

Aronoff (1989) lists six popular misconceptions about remote sensing—especially that done by satellite:

- Satellite-based remote sensing does not have sufficient resolution.
- Satellite data are not sufficiently accurate for practical applications.
- Satellite data are too expensive.
- Remote sensing other than aerial photography is only experimental.
- Remote sensing data are too complicated to use.
- Remote sensing data are not available.

When mapping and classification of large areas are involved, none of these myths are true. Meteorological satellites provide information for specialized natural resource applications. Geosynchronous satellites provide synoptic low-resolution coverage on an hourly basis. The natural resource applications of the U.S. National Oceanic and Atmospheric Administration (NOAA) have increased significantly in the last 5 years. Imagery from the advanced high-resolution radiometer carried aboard the NOAA series of satellites has been used in assessing forest fuel condition and developing national forest cover maps for both the United States and Mexico. Advanced very high resolution radiometer imagery has a nominal resolution of 0.62 miles (1 km) and daily coverage.

The launch of the first earth resources satellite in 1972 added a new dimension to natural resources inventory and monitoring. Today, the U.S. Landsat and French SPOT (Système Probatoire d'Observation de la Terre) satellites provide easily

accessible imagery with global coverage. Circling the earth in polar geosynchronous orbits, the sensors aboard these satellites acquire imagery at a consistent solar time during each daylight pass. Repeat vertical coverage is available from a single satellite on a cycle of approximately 16 days. When multiple satellites in the same series are operating, the orbits are such that the repeat frequency of vertical coverage is proportionally increased.

The current Landsat satellites (4 and 5) carry the Multispectral Scanner (MSS) and the Thematic Mapper (TM). Both of these instruments are mechanical scanners that employ a rotating mirror to acquire data in the cross track direction. The TM has a resolution of 98 feet (30 m) in six bands of reflected energy extending from the blue portion of the spectrum to the middle infrared and an emissive thermal infrared band with a resolution of approximately 394 feet (120 m). TM data have been available since 1982. A panchromatic band with 49 feet (15 m) resolution has been added to the TM to be carried aboard Landsat 6, scheduled for launch in 1993. The panchromatic band imagery can be used to produce ortho-image maps and GIS display backdrops and to enhance the spatial resolution of the 98-foot (30-m) multispectral imagery.

The MSS has a resolution of 262 feet (80 m) in four spectral bands in the green, red, and near-infrared portions of the spectrum. MSS data have been available since 1972. The current Landsat 5 is the last satellite in the series to carry a MSS instrument. Although of significantly lower resolution than the TM, MSS data have been available for more than 20 years, making the data especially suitable for evaluating landscape change.

The French SPOT satellites carry two high resolution visible (HRV) instruments. Unlike the instruments carried aboard the Landsat satellites, the HRV's are solid-state instruments that image the entire swath of the flight path simultaneously. Each of these sensors aboard SPOT 1, 2 (in orbit), and 3 can acquire imagery in the green, red, and near-infrared portions of the spectrum. SPOT 4, scheduled for launch in the middle of the decade, will add a mid-infrared band to the SPOT HRV's. SPOT multispectral imagery has a resolution of 66 feet (20 m). The SPOT HRV's can also be programmed to acquire panchromatic imagery with 30 feet (10 m) resolution. The capability to point these sensors off nadir parallel to the spacecraft ground track permits the acquisition of additional imagery between satellite overpasses and stereo imagery.

For scheduling and archiving, each continuous flight track of earth resources satellites is divided into rectangular data sets or scenes. A full scene of Landsat data is 115 by 115 miles (185 by 185 km), and a full scene of SPOT imagery is 37 by 37 miles (60 by 60 km). Distributors for Landsat and SPOT imagery can supply geocoded digital data suitable for processing with geographic information system (GIS) data. Products available from Spot Image, the U.S. distributor of SPOT imagery, include full scenes, merged data sets covering larger areas, ortho-image quads, and full scene terrain models. Earth Observation Satellite Corporation (EOSAT), the U.S. distributor of Landsat data, can provide data for areas ranging from a full scene covering more than 10,000 square miles (25,890 km^2) to a single USGS quadrangle of coverage on floppy disks. Users can get imagery from either system as digital data or as hardcopy photographic products.

Table A-1 provides broad guidelines to the various wavelengths and applications available using the most common earth-observing satellites.

Satellite imagery may be processed using computer-assisted classification procedures to assign cover classes to the individual pixels in an image or to evaluate change in ground conditions over time. Satellite imagery can be combined with cartographic data to produce image maps. Imagery from the individual bands of satellite imagery can be combined to produce a wide range of products for visual interpretation. The visual perception of the product is influenced by both the bands selected and the assignment of colors to the bands. Color images are produced by assigning primary colors to three selected bands of imagery. The Canadian Centre for Remote Sensing has developed procedures to provide consistent images enhancing boreal, mixed wood, softwood, and leaf-off forest conditions (Ahern and Sirios 1989).

Table A-2 provides examples of various band combinations for TM red, green, and blue image processing configurations.

Various ordering aids are available from SPOT Image and EOSAT. All SPOT data and Landsat imagery acquired since the enactment of the Space Commercialization Act of 1984 are copyrighted. In most cases, the licensing agreement prevents purchasers from sharing the data with other organizations. Landsat data acquired prior to 1984 are not subject to copyright restrictions. Meteorological satellite data are available from NOAA or the USGS Earth Resources Observation System Data Center. Meteorological satellite data are not copyrighted and are available in individual scenes or as seasonal composites. The organizations furnishing satellite imagery will search their archives for data that meet user requirements for location, time period, and maximum cloud cover. Users have the option of scheduling the collection of additional scenes of their area of interest.

Table A-1—Applications of wavelengths available in earth-observing satellites based upon Bain (1991).

Satellite	*Resolution*	*Band*	*Wavelength*	*Application*
AVHRR* (4-channel version)	1–4 km	1	0.55–0.68	Cloud mapping
		2	0.725–1.0	Delineating land/water bodies and melting/nonmelting snow and ice floes
		3	3.55–3.93	Thermal mapping in cloudy areas
		4	10.5–11.3	Mapping sea surface temperatures
		5	11.5–12.5	Removal of radiant energy contribution of water
SPOT Multispectral	20 m	1	0.50–0.59	Green band. Peak vegetation discrimination, vigor assessment
		2	0.61–0.68	Red band. Chlorophyll absorption region aiding in species differentiation, culture identification
		3	0.79–0.89	Near-IR. Vegetation types, vigor and biomass content, water body and soil moisture delineation
Panchromatic	10 m	1	0.51–0.73	Updating features on existing maps and orthophoto maps, monitoring features and detecting change, updating land cover and forest inventory maps
Landsat TM	30 m	1	0.45–0.52	Coastal water mapping; useful for bathymetric mapping of shallow water, soil/vegetation; differentiation, deciduous/conifer differentiation, cultural feature identification
		2	0.52–0.60	Green reflectance by healthy vegetation, useful for vigor assessment, cultural feature identification, discriminating among vegetation types
		3	0.63–0.69	Chlorophyll absorption; useful for plant species differentiation
		4	0.76–0.90	Biomass surveys, delineation of water and vegetation types, assessing vigor and soil moisture
		5	1.55–1.75	Vegetation moisture measurement, snow/cloud differentiation, soil moisture measurement, thin cloud penetration
		6	10.4–12.5	Plant heat stress management, vegetation stress analysis, soil moisture discrimination, thermal mapping applications
		7	2.08–2.35	Hydrothermal mapping, mineral and rock type, vegetation moisture content

*Band combination of 1 and 2 is generally used for vegetation vigor, mapping, and normalized difference vegetation index.

Table A-2—Landsat TM spectral combinations, sequences, appearances, and characteristics based upon Bain (1991).

Band sequence	*Image appearance*	*Characteristics*
2,3,4	Color IR photo	Roads, water bodies, deciduous/conifer differences; high contrast between irrigated and nonirrigated crops
2,4,3	Natural color	Roads; poorer deciduous/conifer contrast than with bands 2,3,4
3,4,5	Natural color	Overall portrayal of vegetation type and condition
3,5,4	Similar to IR	More orange. Vegetation type and condition; more visible roads than with bands 3,4,5
3,4,7	Natural color	Definition of burned areas; revegetation of cut areas visible earlier than with bands 3,4,5; very sensitive to vegetation damage
1,2,3	Natural color	Very high contrast between vegetation and bare areas; very little vegetation information
7,5,3	Similar to IR	Assessing damages caused by fires; burn perimeter red, unburned vegetation green, active fires bright yellow between bands 5 and 7
7,4,3	Similar to IR	Assessing damages caused by fires; burned-over in tones of magenta, burn perimeter and active fires bright red, smoke pale blue, unburned vegetation in tones of green

NOTE—These are only examples. Analysis may use other configurations depending on the classification methods.

Bibliography

Ahern, F.J.; Sirios, J. 1989. Reflectance enhancements for the thematic mapper: an efficient way to produce images of consistently high quality. Photogrammetric Engineering and Remote Sensing 55:61–67.

Aronoff, S. 1982. Classification accuracy: a user approach. Photogrammetric Engineering and Remote Sensing 48(8):1299–1307.

Aronoff, S. 1982. The map accuracy report: a user's view. Photogrammetric Engineering and Remote Sensing 48(8):1309–1312.

Aronoff, S. 1985. The minimum accuracy value as an index of classification accuracy. Photogrammetric Engineering and Remote Sensing 51(1):99–111.

Bailey, R.G.; Pfister, R. D.; Henderson, J.A. 1978. Nature of land and resource classification—a review. Journal of Forestry 76(10):650–655.

Bailey, R.G. 1985. The factor of scale in ecosystem mapping. Environmental Management 9:271–276.

Beckett, P.H.; Burrough, P.A. 1971. The relation between cost and utility in soil survey, IV: comparison of the utilities of soil maps produced by different survey procedures, and to different scales. Journal of Soil Science 22(4):466–480.

Beckett, P.H.; Burrough, P.A. 1971. The relation between cost and utility in soil survey, V: the cost-effectiveness of different soil survey procedures. Journal of Soil Science 22(4):481–489.

Brink, G.E.; Schreuder, H.T. 1992. ONEPHASE: a simulation program to compare 1-phase sampling strategies. Res. Pap. RM-302. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Forest and Range Experiment Station. 8 p.

Brown, P.J. 1982. Multivariate calibration. Journal of the Royal Statistical Society B 3: 287–321.

Congalton, R.G. ed. 1994. Unlocking the puzzle. Proceedings International Symposium on the Spatial Accuracy of Natural Resource Data Bases; 1994 May 16-20; Williamsburg, VA. Bethesda, MD: American Society for Photogrammetry and Remote Sensing. 271 p.

Czaplewski, R. L.; Alig, R. J.; Cost, N. D. 1988. Monitoring land/forest cover using the Kalman filter: a proposal. In: Ek, A.R.; Shifley, S.R.; Burk, T.E., eds. Proceedings, Forest Growth Modeling and Prediction; 1987 August 23–27; Minneapolis, MN. Gen. Tech. Rep. NC-120. Saint Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Forest Experiment Station: 1089–1096.

Czaplewski, R.L. 1990. Kalman filters to update forest cover estimates. In: LaBau, V.J.; Cunia, T., eds. Proceedings, State-of-the-Art Methodology of Forest Inventory; 1989 July 30–August 5; Syracuse, NY. Gen. Tech. Rep. 263. Portland, OR: U.S. Department of Agriculture, Forest Service, Pacific Northwest Forest and Range Experiment Station: 457–465.

Czaplewski, R.L. 1992. Misclassification bias in areal estimates. Photogrammetric Engineering and Remote Sensing 58:189–192.

Czaplewski, R.L. 1994. Variance approximations for assessment of classification accuracy. Res. Pap. RM-316. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Foest and Range Experiment Station. 29 p.

Driscoll, R.; Merkel, D.; Radloff, D.; Snyder, D.; Hagihara, J. 1984. An ecological land classification framework for the United States. Misc. Pub. No. 1439. Washington, DC: U.S. Department of Agriculture, Forest Service. 56 p.

DuPont Corp. Undated. Terminology for DuPont engineering reproduction systems literature. Dupont Corp.: Engineering Reproduction Systems Division. 6 p.

Environmental Systems Research Institute. 1992. Understanding GIS; the Arc/Info method. Revision 6. Redlands, CA: Environmental Systems Research Institute, Inc. 530 p.

Goodchild, M.F.; Duduc, O. 1987. A model of error for choroplethic maps, with applications to geographic information systems. In: Chrisman, N.R., ed. Proceedings, Auto Carto 8; 1987 March 29–April 3; Baltimore, MD. Falls Church, VA: American Society for Photogrammetry and Remote Sensing: 165–174.

Green, E.J. 1987. Empirical Bayes procedures for updating forest inventories. In: Forests, the world and the profession. Proceedings of the Society of American Foresters National Convention; 1986 October 5–8; Birmingham, AL. SAF 87-02. Bethesda, MD: Society of American Foresters: 67–69.

Gregoire, T.G.; Valentine, H.T.; Furnival, G.M. 1987. Importance sampling in forest inventory. In: Forests, the world, and the profession. Proceedings of the Society of American Foresters National Convention; 1986 October 5–8; Birmingham, AL. SAF 87-02. Bethesda, MD: Society of American Foresters: 70–73.

Gregoire, T.G.; Valentine, H.T.; Furnival, G.M. 1993. Estimation of bole surface area and bark volume with Monte Carlo methods. Biometrics 49:653–660.

Haining, R.P. 1990. Spatial data analysis in the social and environmental sciences. Cambridge, UK: Cambridge University Press. 409 p.

Hanson, Russell E. 1979. Preparation of maps for manual digitizing. Res. Inv. Notes BLM-22. Denver, CO: U.S. Department of Interior, Bureau of Land Management, Denver Service Center. 1–8.

Hedges, L.V.; Olkin, I. 1985. Statistical methods for meta-analysis. Orlando, FL: Academic Press. 369 p.

Hinkle, R.E. 1981. Nationwide Forestry Applications Program high altitude imagery user's guide. Houston, TX: U.S. Department of Agriculture, Forest Service, Nationwide Forestry Applications Program. 126 p.

Hopkins, D.H. 1977. Methods of generating land suitability maps: a comparative evaluation. American Institute of Planners Journal 43(4):386–400.

Huff, D. 1954. How to lie with statistics. New York: W.W. Norton and Company, Inc. 142 p.

Jensen, J.R. 1986. Introductory digital image processing: a remote sensing perspective. Englewood Cliffs, NJ: Prentice-Hall. 379 p.

Keates, J.S. 1973. Cartographic design and production. New York: Halsted Press. 240 p.

Keefer, B.J.; Smith, J.L.; Gregoire, T.G. 1991. Modeling and evaluating the effects of stream mode digitizing errors on map variables. Photogrammetric Engineering and Remote Sensing 57(7):957–963.

Laurini, R.; Thompson, D. 1992. Fundamentals of spatial information systems. London: Academic Press. 670 p.

Maling, D.H. 1989. Measurements from maps. Elmsford, NY: Pergamon Press. 577 p.

Mandelbrot, B.B. 1983. The fractal geometry of nature. New York: W.H. Freeman and Company. 468 p.

Mapping Science Committee. 1990. Spatial data needs: the future of the national mapping program. Washington, DC: Board on Earth Sciences and Resources, National Research Council. 78 p.

Martin, F.C. 1985. Using a geographic information system for forest land mapping and management. Photogrammetric Engineering and Remote Sensing 51(11):1753–1760.

Matthews, J.A. 1979. The vegetation of the Storbreen Gletschervorfeld, Jotunheimen, Norway, I: introduction and approaches involving classification. Journal of Biogeography 6:17–47.

Matthews, J.A. 1979. The vegetation of the Storbreen Gletschervorfeld, Jotunheimen, Norway, II: approaches involving ordination and general conclusions. Journal of Biogeography 6:133–167.

Maxim, L.D.; Harrington, L. 1983. Aerial survey design: a systems analytic perspective. Photogrammetric Engineering and Remote Sensing 49(10):1425–1435.

Maxim, L.D.; Harrington, L. 1984. On optimal two-stage cluster sampling for aerial surveys with detection errors. Photogrammetric Engineering and Remote Sensing 50(11):1613–1627.

Maybeck, P.S. 1979. Stochastic models, estimation, and control. Vol. 1. New York: Academic Press. 423 p.

McBratney, A.B.; Webster, R. 1981. The design of optimal sampling schemes for local estimation and mapping of regionalized variables—II. Computers and Geoscience 7(4):335–365.

McBratney, A.B.; Webster, R.; Burgess, T.M. 1981. The design of optimal sampling schemes for local estimation and mapping of regionalized variables—I. Computers and Geoscience 7(4):331–334.

Mitchell, H.; Simmons, L. 1945. The state coordinate systems. Special Pub. 235. Washington, DC: U.S. Department of Commerce, Coast and Geodetic Survey. 62 p.

Monmonier, Mark. 1991. How to lie with maps. Chicago, IL: The University of Chicago Press. 176 p.

Muller, J-C. 1987. The concept of error in cartography. Cartographica 24(2):1–15.

Myhre, R.J; Graham, L.; Sumpter, C. 1991. Airborne videography—a potential tool for resource managers. In: Proceedings, Resource Technology 90; 1990 November 12–15; Washington, DC. Bethesda, MD: American Society for Photogrammetry and Remote Sensing: 591–194.

Prisley, S.P.; Gregoire, T.G.; Smith, J.L. 1989. The mean and variance of area estimates computed in an arc-node geographic information system. Photogrammetric Engineering and Remote Sensing 55(11):1601–1612.

Robinson, A.; Sale, R. 1994. Elements of cartography. 6th edition. Madison, WI: University of Wisconsin. 560 p.

Rosenfield, G.H.; Fitzpatrick-Lins, K.; Ling, H.S. 1982. Sampling for thematic map accuracy testing. Photogrammetric Engineering and Remote Sensing 48(1):131.

Schreuder, H.T.; Wood, G.B. 1986. The choice between design-dependent and model-dependent sampling. Canadian Journal of Forest Research 16:260–265.

Schreuder, H.T.; Gregoire, T.G.; Wood, G.B. 1992. Sampling methods for multiresource inventory. New York: John Wiley & Sons. 446 p.

Schwarz, C., ed. 1989. North American datum of 1983. NOAA Prof. Pap. No. 2. Washington, DC: U.S. Department of Commerce, National Geodetic Survey. 256 p.

Skidmore, A.K.; Turner, B.J. 1990. Assessing the accuracy of resource inventory maps. In: Lund, H. G.; Preto, G., tech. coords. Global natural resource monitoring and assessments: preparing for the twenty-first century. Proceedings of the International Conference and Workshop; 1989 24–30 September; Venice, Italy. Bethesda, MD: American Society for Photogrammetry and Remote Sensing: 524–535.

Smith, J.L.; Logan, J.A.; Gregoire, T.G. 1992. Using aerial photography and geographic information systems to develop databases for pesticide evaluations. Photogrammetric Engineering and Remote Sensing 58(10):1447–1452.

Strahler, A.H.; Woodcock, C.E.; Smith J.A. 1986. On the nature of models in remote sensing. Remote Sensing of Environment 20:121–139.

Thomas, R.W.; Huggett, R.J. 1980. Modeling in geography: a mathematical approach. Totowa, NJ: Barnes and Nobel. 338 p.

U.S. Department of Agriculture, Forest Service. 1985. Geometronics products and services guide. FSM 7144.2. Portland, OR: U.S. Department of Agriculture, Forest Service, Pacific Northwest Region. Misc. pagination.

U.S. Department of Agriculture, Forest Service. 1985. National information requirements project. Washington, DC: U.S. Department of Agriculture, Forest Service. 140 p.

U.S. Department of Agriculture, Forest Service. 1991. National geographic information structure. Washington, DC: U.S. Department of Agriculture, Forest Service, Computer Sciences and Telecommunications. 11 p.

U.S. Department of Interior, Geological Survey. 1987. Map projections—a working model. Prof. Pap. 1395. Washington, DC: U.S. Department of Interior, Geological Survey. 383 p.

Welch, R.; Jordan, T.; Ehlers, M. 1985. Comparative evaluations of the geodetic accuracy and cartographic potential of Landsat-4 and Landsat-5 thematic mapper image data. Photogrammetric Engineering and Remote Sensing 51(11):1799–1812.

Wensel, L.C.; John, H.H. 1969. A statistical procedure for combining different types of sampling units in a forest inventory. Forest Science 15(3):307–317.

Wiant, H.V., Jr.; Wood, G.B.; Gregoire, T.B. 1992. Practical guide for estimating the volume of standing sample tree using either importance or centroid sampling. Forest Ecology and Management 49:333–339.

Wood, D. 1992. The power of maps. New York: The Guilford Press. 248 p.

Woodcock, C.E. 1987. The factor of scale in remote sensing. Remote Sensing of Environment 21:311–332.

Glossary

absolute accuracy The degree of perfection in a value determined through evaluation of all error sources and error propagation (Department of Defense 1981); the difference between an estimate and its parametric (population) or standard (true) value, which can be accounted for by all error sources. Parameters are theoretical values that may or may not exist in reality.

accuracy (1) Freedom from error; conformity to some standard or model. Accuracy relates to the quality of a result and is distinguished from precision, which relates to the quality of the operation by which the result is obtained. (2) The degree of conformity with which horizontal positions and vertical values are represented on a map, chart, or related product in relation to an established standard (Department of Defense 1981). (3) In statistical usage, the closeness of estimates to true values or corresponding population values. An accurate estimator carries little or no *bias* (q.v.). It may or may not be precise.

American Survey Foot Unit of measure used in geodetic surveys in North America. Coordinates in the State Plane Coordinate System are expressed in this unit. Basis of definition is 39.37 inches per International Meter. Distinguished from International Foot, an inch of which contains 25.4 millimeters exactly.

analysis (GIS) As opposed to data manipulation, the derivation of new information bringing together and processing the basic data (polygons, lines, points, labels, etc.) (Aldred 1981).

attribute A qualitative characteristic, usually employed in distinction to a quantitative characteristic (Marriott 1990). Map feature attributes for a road might include surface characteristics, maintenance level, width, and length. In a GIS, attribute is usually analogous to a data element or column in a data base table. By contrast, a *map feature label* might provide only a road number (which in turn might refer a user to another source of information about the road), or a *feature code* (cartographic feature code or feature symbol code) might describe the symbol used to illustrate the road on the map (i.e., feature code 105 is two parallel lines with no interior fill) (USDA Forest Service 1990a).

auxiliary information Supplementary knowledge useful to the conduct of an inventory or mapping project or the interpretation of the project's results (Lund 1986b).

band ratios A method whereby ratios of different spectral bands from the same image or from two registered images are taken to reduce certain effects such as topography and to enhance subtle differences of certain features (Richards 1986).

base map A map constructed from original survey(s) of observable phenomena, not interpreted or analyzed, upon which other information may be placed for purposes of comparison or geographical correlation, or construction of other types of maps (USDA Forest Service 1990a). See *map*.

basic data Data that have not been transformed or interpreted (raw data or edited data). Mapped vegetation would be basic data from which location of old growth could be interpreted (USDA Forest Service 1990a).

bias Generally, an effect that deprives a statistical result of representativeness by systematically distorting it, as distinct from a random error that may distort on any one occasion but balances out on the average (Marriott 1990).

cartography The art and science of expressing graphically, by maps and charts, the known physical features of the Earth or of another celestial body; usually includes the works of man and his varied activities (Department of Defense 1981).

categorical resolution The number of categories in a classification system.

cell In a grid mapping system, a defined geometric shape that stores data or defines an area that is labeled; the smallest addressable unit of space (USDA Forest Service 1990a).

classification The systematic grouping of entities into categories based upon shared characteristics (Lund 1986a).

coefficient of variation The ratio of the standard deviation to the mean multiplied by 100 (Lund and Thomas 1989).

coincidence (mapping) The occurrence of two or more map features in the same location; for example, a road centered upon a section line (USDA Forest Service 1990a).

compatible data Two or more mutually exclusive data sets using the same standards and definitions for purposes of combining (Lund 1986a).

condition The quality or status of an entity. Categories include current, past, desired, actual, potential, and survey.

control (mapping) A system of points with established positions and/or elevations used as fixed references in positioning and correlating map features. *Basic control* implies both horizontal and vertical control, determined in the field and permanently marked or monumented, and required to control subordinate surveys. *Geodetic control* takes into account the size and shape of the Earth, implying a reference spheroid representing the geoid and horizontal- and vertical-control datums. *Ground control* is established by ground surveys, as distinguished from control established by photogrammetric methods. The term usually implies geodetic control or basic control (USDA Forest Service 1990a).

control (photogrammetry) Control established by photogrammetric methods, as opposed to control established by ground surveys (USDA Forest Service 1990a).

control point (1) In photogrammetry, any station in a horizontal and vertical control system identified on a photograph and used for correlating the data shown on that photograph. The term is usually modified to reflect the type or purpose. (2) In inventorying, a point located by ground survey with which a corresponding point on a photograph is matched, as a check, in making mosaics (Department of Defense 1981).

coordinates Linear or angular quantities that designate the position that a point occupies in a given reference frame or system; also used as a general term to designate the particular kind of reference frame or system, such as plane rectangular coordinates or spherical coordinates (Department of Defense 1981). *Plane rectangular coordinates* are used to describe a position on a horizontal plane with respect to a specific origin by means of two distances perpendicular to each other. The merit of a rectangular coordinate system is that positions of points, distances, and directions on it can be computed by the use of plane trigonometry. The *State Plane* and *Universal Transverse Mercator* coordinate systems are plane rectangular systems commonly used to describe locations in a GIS. *Grid coordinates* describe positions within a plane rectangular coordinate system based on, and mathematically adjusted to, a map projection so that geographic positions in terms of latitude and longitude can be readily transformed into plane rectangular coordinates. *Geographic* and *geodetic coordinates* describe a position on the Earth in terms of latitude and longitude (USDA Forest Service 1990a).

corporate data Data that in one form or another are in universal use throughout an organization. Corporate refers in the general sense to the organization that is directly involved in the collection, maintenance, and reporting of the data.

corporate data base A collection of data combined into one body for the purpose of sharing, comparing, and aggregating organizationwide, entered from two or more parallel units within an organization.

corporate geographic information system (GIS) An information system that uses a spatial data base to provide answers to queries of a geographical nature through a variety of manipulations (such as sorting, selective retrieval, calculation, spatial analysis, and modeling) that consist of two or more separately entered themes from two or more parallel units within an organization.

data Information organized for analysis or used as the basis for decisionmaking; numerical information suitable for computer processing, usually in units of information that can be precisely defined. Technically, data are raw facts and figures that are processed into information (Freedman 1983).

data base A generic term used to refer to a collection of computerized files that are stored together.

data dictionary Repository of information (metadata) about the definition, structure, and use of data; it does not contain the actual data (USDA Forest Service 1990a).

datum (1) Any numerical or geometrical quantity or set of such quantities that may serve as a reference or base for other quantities. (2) In geodesy, a datum uniquely defined by five quantities. Latitude, longitude, and geoid height are defined at the datum origin. The adoption of specific values for the geodetic latitude and longitude implies specific deflections of the vertical at the origin. A geodetic azimuth is often cited as a datum parameter, but the azimuth and longitude are precisely related by the Laplace condition, so there is no need to define both. The other two quantities define the reference ellipsoid: the semimajor axis and flattening or the semimajor axis and semiminor axis. (3) In leveling, a surface to which elevations are referred; usually mean sea level, but may be mean low water, mean lower low water, or an arbitrary starting elevation (Department of Defense 1981).

derived map A map that is the result of some analysis of original (or previously derived) data. Whereas a stream map might be made from original observations (ground surveys or from aerial photographs), a land use plan map would be derived from a number of other maps and data sources.

digital classification Employing an algorithm or several algorithms to group pixels of a multispectral image with similar characteristics (Colwell 1983).

digital elevation model (DEM) Digital records of terrain elevations for ground positions at regularly spaced, horizontal intervals. Elevation data are available for USGS 7.5-minute topographic quadrangles and 1° by 2° (1:250,000-scale) maps (USDI Geological Survey 1985).

digital enhancement Data filtering and other mathematical processing, including statistical processing, to manipulate pixel values to produce an image that will accentuate features of interest for visual interpretation (Swain and Davis 1978).

digital line graph (DLG) The digital representation of the planimetric information (line map data) usually portrayed on a map. Large scale DLG's are available for 7.5-minute (1:24,000-scale) and 15-minute (1:62,500-scale in Alaska) topographic quadrangle maps. DLG data from 7.5- and 15-minute topographic quadrangles are stored in the geographic coordinate (latitude/longitude) system (USDI Geological Survey 1985). Other scales are also available. The Federal Geographic Data Committee (1993) describes these products and their availability.

digital terrain model (DTM) A land surface represented in digital form by a series of elevation points with known positions or lists of three-dimensional coordinates (USDA Forest Service 1990a).

digitize (1) To convert from an analog representation of data to a digital one, e.g., to represent a position on a surface by a pair of coordinates with finite resolution. (2) To convert graphics into digital data (usually with a digitizer). This includes deciding which geometrical information should be digitized and stored and which additional alphanumeric information must be input to describe the digitized features and the actual input of this information. This process may be manual, semiautomatic, or automatic. For a GIS, digitizing is particularly concerned with recording of spatial location of geographic phenomena in real-world coordinates, but also includes the entry of any alphabetic or numeric data that describe these phenomena. Usually, digitizing involves keyboard entry and/or on a digitizing table (USDA Forest Service 1990a).

digitizer accuracy The maximum error in any axis between a point's true coordinates and recorded coordinates.

digitizer file The raw source file of digitized data used to define cartographic features, usually including both coordinates and descriptions.

digitizer (general purpose) Any analog-to-digital (abbreviated: A/D) converter.

digitizer (graphic) A device for the conversion of graphics into digital data. It consists of a flat or cylindrical surface to hold the graphics and electronics to sense either certain elements of the graphic at predefined positions or the position of a cursor or stylus. Outputs are signals representing pairs of coordinates and in some cases also signals to indicate the quality of trace features. These signals may be recorded by a data recorder (e.g., a magnetic tape recorder), a procedure called "offline digitizing," or processed directly by a computer ("online digitizing").

digitizer/plotter A device that can be used for both digitizing and plotting. The plotting facility is normally used to check the data immediately after digitizing and to record graphically what has already been digitized.

digitizer precision A product of the resolution of the digitizing table, scanning device, or other equipment used in the process.

discrepancy A difference between results of duplicate or comparable measures of a quantity; the difference in computed values of a quantity obtained by different processes using data from the same survey (Department of Defense 1981).

display An output device that produces a visible representation of the data set for quick visual access (Colwell 1983).

distribution The relative frequency with which different values of a variable occur.

ecosystem management An ecological approach to natural resource management used to achieve multiple-use management of the National Forests and Grasslands. It means that we must weigh and blend the needs of people and environmental values in such a way that the National Forests and Grasslands represent diverse, healthy, productive, and sustainable ecosystems.

ecosystems Biotic communities and their environments existing at any scale, from a rotting leaf or log to the whole Earth.

edge matching Matching of map features that continue beyond the boundary of a given map (or map manuscript) to the same feature on an adjoining manuscript (USDA Forest Service 1990a).

editing (1) The process of checking a map or chart in its various stages of preparation to ensure accuracy, completeness, and correct preparation from and interpretation of the sources used, and to assure legible and precise reproduction. Edits are usually referred to by production phase, such as compilation edit or scribing edit (Department of Defense 1981). (2) The addition, deletion, or modification of data, polygons, lines, points, and associated labels. Editing relates mainly to the correction of errors, but can include updating (Aldred 1981).

elevation Vertical distance from a datum, usually mean sea level, to a point or object on the Earth's surface; not to be confused with altitude, which refers to points or objects above the Earth's surface (Department of Defense 1981).

entity A person, place, or thing independent of others (may be related to others) and having a spatial or physical location. We tend to think of entities as represented by points, lines, or polygons on a map or in a GIS. An entity is mobile. If an entity moves a long distance, it may become remote from related features, but it may still be related. An entity may be created, maintained, modified, observed, moved, or removed.

error (1) The difference between an observed or computed value of a quantity and the ideal or true value of that quantity. (2) Generally classified as one of three types: a *blunder* (mistake), which can be identified and corrected; a *systematic error* (bias), either constant or variable, which must be compensated for; and a *random error*, one of the class of small inaccuracies due to imperfections in equipment, surrounding conditions, or human limitations, or to variation existing naturally in the population being examined (Department of Defense 1981).

estimate The particular value yielded by an estimator in a given set of circumstances (Kendall and Buckland 1971).

estimator The rule or method of estimating a constant of a parent population (Kendall and Buckland 1971). The arithmetic mean is an estimator; its value as calculated from a particular sample is an estimate.

evaluation A determination of the worth, quality, significance, amount, degree, or condition of something by careful appraisal and study (Lund 1986a).

event An occurrence (planned, unplanned, natural, catastrophic, etc.) that changes information about an entity.

existing information Knowledge or data that are currently located somewhere and can be retrieved for use (Lund 1986b).

extrapolation The process of estimating the value of a quantity beyond the limits of known values by assuming that the rate or system of change between the last few known values continues (Department of Defense 1981).

feature See *map feature*.

file A collection of information consisting of records pertaining to a single subject (Spatial Data Transfer Committee 1979).

frame-grabber A computer board that converts a standard analog video signal into a digital raster file. Frame-grabber boards can produce 8-, 16-, or 24-bit raster files, in a gray scale (thermal), color composite, or color RGB separates. The microcomputer interface board accepts a video input signal and passes it to a computer monitor. A program signals the frame-grabber to both freeze and digitize one video frame that is displayed on the monitor. Digitizing a video frame transforms each picture element in the frame to a digital representation. Then, software reads the memory of the board and transfers the image into project file raster objects. Most frame-grabbers are produced for personal computers, but UNIX platforms also can be used for capturing video into raster.

geocoding Transformation or tying-in of digitized coordinates and labels to a map coordinate system (Aldred 1981).

geographic (geographical) Signifying basic relationship to the Earth considered as a globe-shaped body. The term geographic is applied alike to data based on the geoid and on spheroids (Department of Defense 1981).

geographic data Information that describes characteristics of the Earth, including its natural and cultural features. Geographic data have either spatial (or locational) or attribute (or descriptive) components, or both (USDA Forest Service 1990a). See *spatial data*, *attribute*.

geographic information system (GIS) A specialized form of data base management system that can be used to enter, edit, manage, manipulate, analyze, query, and display both graphic and tabular data; handles both spatial and attribute data and allows the user to work with these data to create summaries and display spatial relationships (USDA Forest Service 1990a). In this primer, we refer to GIS's that operate on a computer.

geographic position The position of a point on the surface of the Earth expressed in terms of latitude and longitude, either geodetic or astronomic (Department of Defense 1981).

geographically referenced (georeferenced) The condition of data for which positional information is available, enabling the geographical position of the data to be established and communicated (Haddon 1988).

geometronics The art and science of recording, measuring, interpreting, handling, and displaying information about the Earth and its resources; combines the fields of cartography, remote sensing, geodesy, and photogrammetry (USDA Forest Service 1990a).

global positioning system (GPS) A navigation and positioning system with which the three-dimensional geodetic position and the velocity of a user at a point on or near the Earth can be determined in real time. The system consists of a constellation of satellites that broadcast on a pair of ultrastable frequencies. The user's receiver tracks the satellites from any location at any time, thus establishing position and velocity (Department of Defense 1981).

graphic Any and all products of cartographic and photogrammetric art. A graphic may be a map, chart, mosaic, or even a filmstrip produced using cartographic techniques (Department of Defense 1981).

graticule, map A series of straight or curved lines intersecting at right angles, representing latitudes and longitudes on a map or chart.

grid (1) Two sets of straight, parallel lines intersecting at right angles and forming cells; superimposed on maps, charts, and other similar representations of the Earth's surface in an accurate and consistent manner to permit identification of ground locations with respect to other locations and the computation of direction and distance to other points (also called reference grid). (2) A term used in giving the location of a geographic point by grid coordinates (Department of Defense 1981). (3) A process by which a vector map is converted into a grid map; analogous t o tilting, or regular tessellation.

grid map A raster-based data structure wherein space is divided into two-dimensional cells of equal size and regular shape arranged in columns and rows. The attributes of each cell represent the location (row and column) and information about the value or the nature of the geographical feature represented.

ground truth Data and observations on the Earth's surface normally to quantify simultaneously recorded remote sensing imagery (Slama 1980).

image A graphic representation of an object or objects, typically produced by an optical or electronic device, in which the appearance of the object(s) is reproduced as perceived by normal binocular vision. An image can be graphically reproduced on a photographic medium or upon an electronic display device, and likewise may be stored physically on photographic media or logically in a digital electronic file. An image captured, stored or displayed electronically is generally processed as a raster, in which the image is broken up into cells of equal size and shape arranged in columns and rows, called pixels, for which optical characteristics are generalized, stored, and displayed.

imagery Collectively, the representations of objects reproduced electronically or by optical means on film, electronic display devices, or other media (Department of Defense 1981).

information Knowledge derived from study, experience, or instruction.

information needs analysis (assessment) (INA) A definable process that documents what questions need answers when, at what cost, and with what reliability. The purpose of an INA is to identify an organization's requirements for the least quantity of information of the highest quality in the most timely manner (Hoekstra 1982).

integrated inventory An inventory or group of inventories designed to meet multilocation, multidecision level, multiresource, or monitoring needs (Lund 1986a).

interactive The ability of the machine or operator to communicate on a real-time or continuing basis to solve problems (Aldred 1981).

interpolate To determine intermediate values between given fixed values. As applied to logical contouring, to interpolate is to ratio vertical distances between given spot evaluations (Department of Defense 1981).

inventory To account quantitatively for goods on hand or to provide a descriptive list of articles giving, at a minimum, the quantity or quality of each (Lund and Thomas 1989).

inventory (survey) unit The land unit containing the population for which information will be summarized and analyzed (Lund and Thomas 1989).

label Alphanumeric data, textual data, or a symbol that describes a polygon, line, or point. Sometimes referred to as attribute label, type code, or descriptor (Aldred 1981). See also *attribute* ("map feature label").

Lambert conformal conic map projection A conformal map projection on which all geographic meridians (longitude) are represented by straight lines that meet in a common point outside the limits of the map, and the geographic parallels (latitude) are represented by a series of arcs of circles having this common point for a center. Meridians and parallels intersect at right angles, and angles on the Earth are correctly represented on the projection. This projection may have one standard parallel along which the scale is held exact; or there may be two such standard parallels, both maintaining exact scale. At any point on the map, the scale is the same in every direction. It changes along the meridians and is constant along each parallel. Where there are two standard parallels, the scale between those parallels is too small; beyond them, too large. Also called Lambert conformal map projection (Department of Defense 1981).

land cover That which overlays or currently covers the ground, especially vegetation, water bodies, or structures. Barren land is also considered "land cover," although technically it is lack of cover. The term land cover can be thought of as applying to the setting in which action (one or more different land uses) takes place (USDA Forest Service 1989a).

land use The predominant purpose for which an area is employed (USDA Forest Service 1989a).

Landsat imagery Images of the Earth's surface prepared from data sensed and transmitted to receiving stations by the Landsat satellite.

latitude (1) In general, a linear or angular distance measured north or south of the equator on a sphere or spheroid. (2) In plane surveying, the perpendicular distance in a horizontal plane of a point from an east-west axis of reference (Department of Defense 1981).

layer A physical or digital separation of geographic information by theme. (1) In cartography, often a physical product such as a drafted overlay, a film negative, or scribed art work, usually representing information within a single or related themes. (2) In digital applications, often a computer map file. Like the cartographic separation, each computer map file is likely to contain features within a common theme. For example, transportation features (roads and trails) might comprise a layer in cartographic medium or a digital file (USDA Forest Service 1990a).

legend A description, explanation, table of symbols, or other information printed on a map, overlay, or chart to provide a better understanding and interpretation of it (Slama 1980).

line (1) In a GIS, a one-dimensional defined object having a length and direction, and connecting at least two points. Examples of geographic phenomena symbolized as lines on maps are roads, railroads, streams, and telecommunications lines (USDA Forest Service 1990a). (2) Mathematically defined, an infinite length in both directions. A ray extends from a point to an infinite length in one direction, and a line segment is of finite length.

longitude A linear or angular distance measured east or west from a reference meridian (usually Greenwich) on a sphere or spheroid (Department of Defense 1981).

manuscript The final compilation of all information for map construction or digitizing. For digitizing or complex map construction, some of the information may be on one or more overlays (USDA Forest Service 1990a). See *map*.

map (1) A graphic representation, usually on a plane surface and at an established scale and/or projection, of all or a portion of the Earth, showing the relative size and position of natural and artificial features. The features are positioned relative to a coordinate reference system. A map may emphasize, generalize, or omit the representation of certain features to satisfy specific requirements. Maps are frequently categorized and referred to according to the primary type of information that they are designed to convey, to distinguish them from maps of other types. A *topographic map* represents the horizontal and vertical positions of the features represented; it is distinguished from a planimetric map by the addition of relief in measurable form. A topographic map shows mountains, valleys, and plains, and, in the case of hydrographic charts, symbols and numbers to show depths in water bodies. A *contour map* is a topographic map that portrays relief by means of contour lines. A *planimetric (line) map* presents only the horizontal positions for the features represented; it is distinguished

from a topographic map by the omission of relief in measurable form. A *base map* shows certain fundamental information used as a base upon which additional data of a specialized nature are compiled. A *hydrographic map* shows a portion of the waters of the Earth, including shorelines, shoreline and underwater topography, and as much of the topography of the surrounding country as is necessary for the purpose intended. A *map manuscript* is the original drawing of a map as compiled or constructed from various data (such as ground surveys, photographs, or other source materials and data). A *thematic map* concentrates on the spatial relationships of a single attribute or subject, or the relationships among several. The objective is to portray the form or structure of a distribution, that is, the character of the whole as consisting of the interrelation of the parts. Just because a map deals largely with a single subject does not necessarily mean that it is a thematic map. Thematic maps generally employ symbolization that focuses attention on the structure of the distribution, and accuracy is less a matter of positional accuracy than the truthfulness of the portrayal of the distribution's basic structural character (USDA Forest Service 1990a and Department of Defense 1981). (2) To prepare a map or engage in a mapping operation (Department of Defense 1981).

map feature The map representation of a real-world phenomenon or entity, such as a well, town, road, boundary, swamp, or timber stand (USDA Forest Service 1990a).

map projection A systematic drawing of lines on a plane surface to represent the parallels of latitude and the meridians of longitude of the Earth or a section of the Earth, with intention of minimizing distortion in area, shape, distance, and direction (USDA Forest Service 1990a and Department of Defense 1981). A map projection may be established by analytical computation or may be constructed geometrically. A map projection is frequently referred to as a "projection," but the complete term should be used unless the context clearly indicates the meaning (Department of Defense 1981). Commonly used map projections include the Lambert conformal conic and Transverse Mercator.

mapping The identification of selected features, the determination of their boundaries or locations, and the delineation of those boundaries or locations on a suitable base using predefined criteria.

mean The expectation of a variable, usually the *arithmetic mean* of a sample of values. Other means include the *geometric mean* and *quadratic mean.* The mean is the average, however defined, of a group of values.

merge In image processing, the reduction of number of labels and polygons after dissolving lines during reclassification (Aldred 1981).

metadata Data about data, e.g., source, accuracy, and age.

monitoring The collection of serial data to detect changes or evaluate trends as well as to understand how a system functions (Lund 1986a).

Multispectral Scanner (MSS) (1) A remote-sensing device capable of recording data in the ultraviolet and visible portions of the electromagnetic spectrum, as well as the infrared (Department of Defense 1981). (2) As used in the Landsat program, a scanner system that uses an oscillating mirror and an array of six fiber-optic detectors in each of four spectral bands from 0.5 to 1.1 um. The mirror sweeps from side to side in 185-km swaths, transmitting incoming energy to the detector array, which sequentially outputs brightness values (signal strengths) for successive pixels. Image resolution is approximately 80 meters (ground dimension of a pixel).

Mylar Registered trademark for polyester film manufactured by DuPont Corporation. Mylar has broad applications as electrical insulation, magnetic tape base, packaging materials, balloons, and other uses, in addition to its use as photographic and drafting media (USDA Forest Service 1990a).

noncorporate data Data specific to certain subunits or locations within an organization, but not generally accessed by other levels or parallel subunits.

normal distribution function A mathematical function describing the behavior of one-dimensional random errors (Department of Defense 1981). Its parameters are the mean and variance of the distribution. The importance of the normal distribution to sampling and statistics lies in the fact that, for a population having an error distribution of any form so long as its second moment exists, the distribution of sample arithmetic means approaches a normal distribution with increasing sample size. The bivariate and multivariate normal distributions may describe random errors of two or more dimensions.

optical scanner Device that reads graphic images by detection of reflected or transmitted light and creates a digital file representing the image. A *multispectral scanner* (q.v.) measures reflected and transmitted energy across bands of the electromagnetic spectrum beyond that of visible light (USDA Forest Service 1990a). See *remote sensing*.

ortho-image map A photo map made from an assembly of orthophotographs or imagery (Department of Defense 1981).

orthophoto A photograph having the properties of an orthographic projection, that is, the image is transformed to appear as though viewed from a right angle to the image plane. It is derived from a conventional perspective photograph by simple or differential rectification so that image displacements caused by camera tilt and relief of terrain are removed (USDA Forest Service 1990a).

orthophoto quad An orthophoto, derived from aerial photography, positioned or mosaicked to correspond to coverage of a 7.5-minute quadrangle. Thus, the features visible on an orthophoto quad should align with the analogous features on a corresponding quad map. If constructed to meet accuracy standards suitable for the data, it can be used as a base map for manuscript preparation or digitizing (USDA Forest Service 1990a).

overlay (1) A map of a particular subject or theme, portrayed on a transparent or translucent medium, which, when registered to a base map, allows observation and measurement of relationships between the overlay theme and features portrayed on the base map. (2) To combine in an automated process two or more map themes for the same land area to create a new map based on a combination of the original maps. Depending on the software and the operation selected, the result may be only a graphic composite of the images or a logical or arithmetic combination of the themes to produce a new product reflecting relationships between the themes (USDA Forest Service 1990a).

overshoot A cartographic error in which a line extends past its end point. An overshoot is most noticeable when a line crosses another line at which it is supposed to end.

parameter (1) In general, any quantity of a problem that is not an independent variable. More specifically, the term is often used to distinguish from dependent variables quantities that may be assigned more or less arbitrary values for purposes of the problem at hand (Department of Defense 1981). (2) In statistics, a characteristic of a population, often unknown, as compared to a statistic, which is a characteristic of a sample. Statistics often serve as estimators of population parameters.

photo map A reproduction of a photograph or photomosaic upon which the grid lines, marginal data, contours, place names, boundaries, and other data may be added (Department of Defense 1981).

photogrammetry (1) The art, science, and technology of obtaining reliable information about physical objects and the environment through processes of recording, measuring, and interpreting images and patterns of electromagnetic radiant energy and other phenomena (Richards 1986). (2) The preparation of charts and maps from aerial photographs using stereoscopic equipment and methods (Department of Defense 1981). See also *stereo plotter*.

photographic (image) interpretation The examination of photographic images for the purpose of identifying objects and deducing their significance; also called photointerpretation (Department of Defense 1981).

pixel The smallest, most elementary areal constituent of a raster image (also called a Resolution Cell) (Haddon 1988).

platform, remote sensing The vehicle that holds a sensor. It is usually a satellite, but may be an airplane or a helicopter. Sensors can be mounted on tripods for certain uses, such as examining electromagnetic radiation from various types of vegetation (Department of Defense 1981).

point An object that has no dimension, but has geometric location specified by a set of coordinates. Though all geographic phenomena have dimension, their expression on a map as a point symbol is determined by scale. Examples of geographic phenomena usually symbolized as points on large-scale maps are

wells, weather stations, and navigational lights. An airport that appears as a polygon outline of actual runways on a large-scale map might also be shown as a point symbol on a small-scale map (USDA Forest Service 1990a).

polygon A closed plane figure bounded by three or more line segments. A single timber stand delineated on a map or overlay is an example of a polygon. A stream of digitized points approximating the delineation (perimeter) of an area on a map, polygons are often comprised of line segments or arcs that join at nodes to produce a multisided figure (Aldred 1981).

population The aggregate or collection of unit values that contains all of the unit values; it is defined by its members. A subpopulation may carry a secondary definition that characterizes an identifiable subportion of a population.

positional accuracy In cartography, a term used in evaluating the overall reliability of the positions of cartographic features on a map or chart relative to their true position or to an established standard (Department of Defense 1981).

precision (1) The degree of refinement in the performance of an operation or the degree of perfection in the instruments and methods used when making the measurements. Precision relates to the quality of the operation by which a result is obtained and is distinguished from accuracy, which relates to the quality of the result (Department of Defense 1981). (2) In statistics, precision describes the degree to which sample observations or sample-based estimates tend to cluster about their own mean. If the sample is biased so that the expected value of the sample statistic is not the corresponding population parameter, then the statistic may be inaccurate while being highly precise. Conversely, an unbiased estimator is accurate, though it may be very imprecise.

primary base series (PBS) USDA Forest Service 1:24,000-scale quadrangle maps. These maps are composed of layers, including hydrology, transportation, boundaries, topography, land net, land status, and culture (buildings, campgrounds, fences, reservoirs, etc.).

processing The manipulation of data by means of a computer or other device (Slama 1980).

projection See *map projection*.

Public Land Survey System (PLSS) The township, range, and section grid established in the United States by the Land Ordinance of 1785. This is the legal survey system used to subdivide lands in most of the United States, outside of the original 13 colonies on the east coast. It is a rectangular survey system using township squares that are 6 miles on a side (36 square miles) as the basic survey unit. The location of townships is controlled by baselines and meridians running parallel to latitude and longitude lines. Townships are defined by range lines running parallel (north and south) to meridians and township lines running parallel (east and west) to baselines.

puck A handheld device moved freely around on a digitizer surface (similar to an ice-hockey puck on ice); used to indicate the location to be digitized by means of a crosshair or other reference mark (also called a cursor).

quadrangle A rectangular, or nearly so, area covered by a map or plat, usually bounded by given meridians of longitude and parallels of latitude; also called quad or quadrangle map (Department of Defense 1981).

raster A division of 2-dimensional space into regular polygons (usually rectangular) ordered in line scan form, that is, one row followed by another. Data that comprises a set of pixels usually arranged on rectangular grid centers (Richards 1986).

rectification In photogrammetry, the process of projecting a tilted or oblique photograph onto a horizontal reference plane. Although the process is applied principally to aerial photographs, it may also be applied to the correction of map deformation (Department of Defense 1981).

registration In cartography, the maintenance of relative position between features on various layers of information. *Physical registration* refers to the maintenance of relative position between various sheets of map material, such as drafting film or overlays, to a base map. Two methods of physical registration are commonly used: visual and mechanical. *Visual registration* involves visible graphic marks, which, when aligned, ensure proper relative position between layers. *Mechanical* or *pin register systems* use pins that are inserted in matched holes in the materials. *Geodetic registration* refers to the establishment of absolute ground coordinates to specific map (control) locations to establish absolute Earth positioning for the remainder of the map (USDA Forest Service 1990a).

regression analysis (1) An analysis that investigates how one variable is related to another by providing an equation that allows the use of the known value of one or more variables to estimate the unknown value of the remaining variables. (2) A method of deriving an estimator for a characteristic of a population that is difficult or expensive to measure, based upon its relationship to other more easily measured variables. The characteristic predicted is the dependent variable; those that serve as predictors are the independent variables. They are related by a mathematical expression called the *regression model*, the parameters of which are estimated by regression analysis. Usually, regression analysis is a least squares procedure.

relational data base A structured set of data, containing sets of records or rules so that relations between different entities and attributes can be used for data access and transformation (USDA Forest Service 1990a).

relative accuracy In general, an evaluation of the random errors in determining the positional orientation (e.g., distance and azimuth) of one point or feature with respect to another (Department of Defense 1981).

remote sensing The measurement or acquisition of information of some property of an object or phenomenon by a recording device that is not in physical or intimate contact with the object or phenomenon under study; sometimes restricted to the practice of data collection in the wavelengths from ultraviolet to radio regions (Department of Defense 1981).

report A document that displays tabular plus supportive information such as title, date produced, and footnote information. Reports can be either displayed on terminals or produced on paper.

representative fraction (RF) The scale of a map or chart expressed as a fraction or ratio. Relates unit distance on the map to distance measured in the same unit on the ground; also called fractional scale or natural scale (Department of Defense 1981).

residual error The difference between any value of a quantity in a series of observations, corrected for known systematic errors, and the value of the quantity obtained from the combination or adjustment of that series; frequently used as the difference between an observed value and the mean of all observed values of a statistically valid set (Department of Defense 1981).

resolution The minimum distance between two adjacent features, or the minimum size of a feature, that can be detected by a remote sensory system (Department of Defense 1981); expressed as the spacing measured on the image in line-space pairs per unit distance of the most closely spaced lines that can be distinguished. The term is also used to coincide with the dimension, or aerial extent, of a pixel. Map resolution may refer to a "minimum mapping unit" (only shows lakes of 5 acres [2 ha] or greater) or the accuracy at which a given map scale can depict the location and shape of map features.

resource data In a GIS, geographic data describing natural resource phenomena by both attribute and position (USDA Forest Service 1990a).

resource inventory The collection of data for description and analysis of the status, quantity, quality, or productivity of a resource. Such inventories usually include descriptive data, numeric data, and, at times, maps showing the extent of the inventory unit, the resources, and the location of sample units (Lund and Thomas 1989).

sample A subset of one or more of the sample units into which the population is divided that is selected to represent the population and examined to obtain estimates of population characteristics.

sample plot A sampling unit or element of known area and shape, such as a 0.2-ha rectangular plot.

sample size The number of sampling units included in a sample.

sampling The selection of sample units from a population and the measurements and/or recording of information contained therein to obtain estimates of population characteristics.

sampling (inventory) design The specification of an allocation or configuration of sampling units and the method used to determine which sampling units will be measured.

sampling error That part of the difference between a population value and an estimate thereof derived from a random sample (Kendall and Buckland 1971).

sampling frame The complete aggregate or list of sampling units from which the samples will be drawn; may be a real list for finite population sampling or a theoretical construct for sampling from infinite or noncountable populations.

sampling intensity The sample size relative to the population size; often expressed as the number of sampling units drawn divided by the total number in the population, when the latter is known. When sampling units are areal in nature, as are fixed area plots, sampling intensity may be expressed as the number of sampling units drawn per unit of the area being sampled.

sampling unit One of the specified parts into which the population has been divided for sampling purposes (Kendall and Buckland 1971). A sampling unit may be of largely natural definition (e.g., a person, plant, or city) or defined by the sampler (e.g., sample plots, plot clusters, or strips). Each sample unit commonly consists of only one sample element, which may be a sample plot, tree, or shrub.

scanner, optical A device which reads graphic images by detection of reflected or transmitted light, and creates a digital file representing the image. A multispectral scanner such as those used on earth observations satellites, measures reflected and transmitted energy across bands of the electromagnetic spectrum including and beyond that of visible light, and record spectral (i.e. color) characteristics of the energy in addition to its intensity. Bi-level, or black-and-white output scanners only sense the intensity of received energy and record only a binary (on or off) condition for each pixel, as determined by a threshold parameter. Such scanners are used to convert map and textural graphics into computer acceptable files.

scale The relation between the distance on a photograph or a map to its corresponding distance on the ground. Scale may be expressed as a ratio (1:24,000); a representative fraction (1/24,000); or an equivalence (1 in = 2,000 ft). The scale of a photograph varies from point to point due to displacements caused by tilt and relief, but is usually taken as f/H, where f is the focal length of the camera and H is the height of the camera above mean ground elevation (USDA Forest Service 1990a).

scientifically valid design A sampling and estimation scheme in which sample units are defined and chosen according to the accepted science for that particular resource or sector and reflect a statistical basis for sampling. See *statistically valid design*.

secondary base series The Forest Service series of base maps, at 1:126,720 scale, generally covering a national forest, grassland, or purchase unit. Forest Service visitors' maps are made from secondary base series maps (USDA Forest Service 1990a).

sensitivity A measure of how fine measurements or interpretations need to be to distinguish between classes. Some categories in a classification system are likely to be easy to distinguish between and others difficult. For example, in forest inventories, it is very easy to distinguish between classes that have drastic differences in the amount of vegetation present, whereas it is more difficult to distinguish between categories with similar vegetation amounts but different structural characteristics. The sensitivity required to distinguish between classes is also affected by contextual considerations, such as the degree of contrast between an object and the objects that surround it.

six-parameter affine solution A transformation process involving six unknown parameters, including rotation, nonperpendicularity of the axes, two scale changes, and two translators. The six-parameter affine solution is often called the two-dimensional affine transformation. The difference is only slight; the two-dimensional transformation is expanded to six parameters, allowing for two scale factors instead of one and nonperpendicularity (or affinity) between the two axes of the system to be rotated.

software A set of computer programs, procedures, and associated documentation (if any) concerned with the operation of a data processing system (Slama 1980).

spatial data Data describing location or position in space; in this primer, used to distinguish the locational component of geographic data from the attribute or other descriptive component. In a GIS, spatial data are generally considered to be the specific location identifiers or coordinates used to describe location (USDA Forest Service 1990a). See *coordinates.*

spatial data base A collection of interrelated, geographically referenced data stored without unnecessary redundancy to serve multiple applications as part of a geographic information system (Haddon 1988).

spatial resolution The smallest discernible spatial unit. For photographic imaging systems (film/camera combinations), the spatial resolution is usually expressed as the maximum number of line-space pairs per unit of distance area that can be clearly detected on a photographic product. For digital imagery, spatial resolution is usually expressed as pixel size. For field surveys, the spatial resolution may be expressed as the density of sample points.

specifications The rules, regulations, symbology, and a comprehensive set of standards that have been established for a particular map or chart series or scale group. Specifications vary with the scale and the purpose of the graphic (Department of Defense 1981).

standard deviation The positive square root of the variance; sometimes called RMS for "root mean square." Standard deviation usually refers to variation among observations (Marriott 1990).

standardization (1) The comparison of an instrument or device with a standard to determine the value of the instrument or device in terms of an adopted unit (Department of Defense 1981). (2) The act of bringing items into conformity with quantitative or qualitative criteria commonly used and accepted as authoritative (Lund 1986a). (3) A linear transformation of a random variable that consists of subtracting its mean and dividing the result by its standard deviation. The transformed random variable then has a mean of zero and a standard deviation of unity.

State Plane Coordinate System (SPCS) The plane-rectangular coordinate systems (one for each State in the United States) established by the National Geodetic Survey for use in defining positions of geodetic stations in terms of plane-rectangular (x and y) coordinates. Also called State System of Plane Coordinates (Department of Defense 1981). Each State system consists of one or more zones. The grid coordinates for each zone are based on, and mathematically adjusted to, a map projection. The *Lambert conformal conic map projection* (q.v.) with two standard parallels is used for States of predominantly east-west extent. The Transverse Mercator projection (see *Transverse Mercator grid*) is used for States of predominantly north-south extent. The north and east directions are taken as positive, and to avoid the use of negative coordinates, the origin of each zone is established at a point to the southwest of the land area intended to be served by the zone. The unit of measure in the SPCS is the American Survey Foot (USDA Forest Service 1990a).

statistically valid design A scheme in which sample units are chosen from the population of interest, utilize objective observations, and permit the calculation of sampling error (Lund 1986a).

stereo plotter A device for constructing an orthographic projection or obtaining spatial information in the form of coordinates by observation of a pair of overlapping photographs (USDA Forest Service 1990a).

stratification The division of an inventory unit into more homogeneous subunits to improve the efficiency of the inventory (Lund and Thomas 1989). Stratification may also be used to segregate information by meaningful subdivisions of the population. Even when this is its primary justification, some increase in sampling efficiency is a usual result.

stratum Any division of the population for which a separate estimate is desired (Kendall and Buckland 1971).

systematic error An error that occurs with the same sign, and often with a similar magnitude, in a number of consecutive or otherwise related observations. For example, when a base is measured with a wrongly calibrated tape, there will be systematic errors. In addition, random errors will occur. Repetition does little or nothing to reduce the ill effect of systematic errors, which are a most undesirable feature of any set of observations. Much of the care in making observations is directed toward eliminating or correcting systematic errors (Department of Defense 1981). Systematic error is *bias* (q.v.).

tabular data Numeric and character data arranged in rows and columns used to provide descriptive information about graphic features.

temporal resolution The time frame over which successive measurements are taken. Temporal resolution is important to consider when we attempt to inventory and monitor dynamic systems, particularly when it is necessary to integrate data collected over significantly different time periods.

Thematic Mapper (TM) A scanner having more spectral, radiometric, and geometric sensitivity than its predecessors, part of the payload of Landsat satellites since Landsat 4 (Haddon 1988).

theme The subject matter of a map or data layer containing information regarding related phenomena. For example, the theme hydrography might include river and other stream locations, lake and reservoir boundaries, springs, and gauge station locations. Though the data types and map symbols may vary, they are considered to be within a common theme. Compare to *layer* (q.v.) (USDA Forest Service 1990a).

tie A survey connection from a point of known position to a point whose position is desired. A tie is made to determine the position of a supplementary point whose position is desired for mapping or reference purposes, or to close a survey on a previously determined point. To "tie in" is to make such a connection (Department of Defense 1981).

tone Each distinguishable shade variation from black to white on imagery (Department of Defense 1981).

topographic map A map that presents the vertical position of features in measurable form as well as their horizontal positions (Department of Defense 1981).

transformation (1) In photogrammetry, the process of projecting a photograph (mathematically, graphically, or photographically) from its plane onto another plane by translation, rotation, and/or scale change. The projection is made onto a plane determined by the angular relations of the camera axes and not necessarily onto a horizontal plane. (2) In surveying, the computational process of converting a position from *Universal Transverse Mercator* (q.v.) or other grid coordinates to geodetic, and vice versa, and from one datum and ellipsoid to another using datum shift constants and ellipsoid parameters (Department of Defense 1981). (3) In statistics, a mathematical operation performed on a variable to obtain another variable more amenable to statistical analysis. Commonly used transformation functions are the logarithm, the square, the reciprocal, and the sine, cosine, and tangent for topographic characteristics.

transmittance The ability of a substance to transmit energy, expressed as the ratio of the energy transmitted through a body to that incident upon it.

Transverse Mercator grid An informal designation for a State coordinate system based on a Transverse Mercator map projection; also called Gauss-Kruger grid (Department of Defense 1981).

trend The measure of change in variables, such as growth or ecological status, observed over time.

type map A map or overlay showing the distribution of various features, and specific classes of features such as soil, vegetation, or site, throughout a given area (Ford-Robertson 1971).

U.S. National Map Accuracy Standards (NMAS) (1) Horizontal accuracy: for maps at publication scales larger than 1:20,000, 90 percent of all well-defined features, with the exception of those unavoidably displaced by exaggerated symbolization, will be located with 1/30 inch (0.85 mm) of their geographic positions as referred to the map projection; for maps at publication scales of 1:20,000 or smaller, 1/50 inch (0.50 mm). (2) Vertical accuracy: 90 percent of all contours and elevations interpolated from contours will be accurate within one-half of the basic contour interval. Discrepancies in the accuracy of contours and elevations beyond this tolerance may be decreased by assuming a horizontal displacement within 1/50 inch (0.50 mm). Also called map accuracy standards; national map accuracy standards (Department of Defense 1981).

undershoot A cartographic error in which a line does not extend to its end point. An undershoot is most noticeable where lines are supposed to meet but don't.

Universal Transverse Mercator (UTM) projection The ellipsoidal form of the Transverse Mercator to which specific parameters, such as central meridians, have been applied. It is a widely used map projection employing a series of identical zones, each covering 6 degrees of longitude and each oriented to a specific central meridian. The UTM projection is characterized by its property of conformity, the preservation of constant scale along lines approximately parallel to the central meridian, and a maximum scale distortion of 1 part to 1,000. Each geographic location in the UTM projection is given x and y coordinates, in meters. The UTM is one of the projection options offered by NASA for Landsat data and is the most common projection used for Landsat image maps.

Universal Transverse Mercator (UTM) grid system A grid system originally adopted by the U.S. Army in 1947 for designating rectangular coordinates on large-scale military maps of the entire world, later adapted for use in civilian mapping. It provides coordinate locations for all points on the globe between 84° N. latitude and 80° S. latitude, based on a series of maps in the Universal Transverse Mercator projection. The Earth is divided into 60 zones each generally 6° wide in longitude. Zones are numbered from 1 to 60 proceeding east from the 180th meridian from Greenwich, with minor exceptions. For example, Washington, DC, is in UTM grid zone 18. Unit of measure is meters.

update To address change within a data acquisition (inventory or mapping) cycle. Updating is the procedure of modifying a portion of an existing data set through sampling, mechanical, or modeling procedures to the present time (Lund 1986a).

variable A characteristic that may vary from sample unit to sample unit, such as tree height, diameter, species, or sex.

variance (1) The square of the standard error; defined as the limit, as the number of observations becomes infinitely large, of the sum of the squares of the residuals divided by *n*: the mean of the mean of the squares of errors (Department of Defense 1981). (2) The measure of dispersion of individual unit values about their mean (Kendall and Buckland 1971). (3) The expected value of the squared difference between the value of a random variable and its mean. The variance is the second moment of a distribution about its mean.

variation The dispersion of values of a random variable about its mean.

vector A file of points such that magnitudes and direction can be drawn from point to point (in principle) to reconstruct line segments on a display or plotter (Richards 1986).

vector data In a GIS, data composed of x-y coordinate representations of locations on the Earth; takes the form of single points, strings of points (lines), or closed lines (polygons).

vegetation cover map A map or overlay prepared to show the location and general vegetation composition of the various strata comprising an inventory unit.